5-2005

Hand-Held Calculators And Mathematics Achievement: What the 1996 National Assessment Of Educational Progress Eighth-Grade Mathematics Exam Scores Tell Us

Kenneth L. Wareham
Utah State University

HAND-HELD CALCULATORS AND MATHEMATICS ACHIEVEMENT: WHAT THE 1996 NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS EIGHTH-GRADE MATHEMATICS EXAM SCORES TELL US

by

Kenneth L. Wareham

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Psychology

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

2005

ABSTRACT

Hand-Held Calculators And Mathematics Achievement: What the 1996 National Assessment Of Educational Progress Eighth-Grade Mathematics Exam Scores Tell Us

by

Kenneth L. Wareham, Doctor of Philosophy

Utah State University, 2005

Major Professor: Dr. George Julnes
Department: Psychology

The purpose of this study was to analyze the 1996 National Assessment of Educational Progress data to identify the relationship between calculator use and student performance on the National Assessment of Educational Progress Mathematics Assessment. This general purpose includes several sub issues. In addition to being interested in the overall relationship between use and National Assessment of Educational Progress achievement (including the effort to control for spurious factors), this study examined the contextual factors that moderate the impact of calculator use. Similarly, it analyzed the relationship between calculator use and student performance on calculator-allowed and calculator-restricted items, as well as the ability of students to recognize whether the use of a calculator was appropriate when responding to a math problem.

Findings indicate that significant differences in achievement exist between students who regularly use calculators and those who do not use calculators. Even when controlling for various contextual factors that moderated this relationship (e.g., gender, socioeconomic status, parents' level of education, students' National Assessment of Educational Progress achievement level), it was found that the more frequently students use a calculator the higher their scores tend to be. The results also show that when not allowed to use calculators, the more frequent calculator users continue to score higher than those who do not use calculators. Finally, using calculators does not automatically equate to calculator dependence, and, in fact, the more often students use a calculator the more adept they are at applying it properly and withholding it when inappropriate.

Based on the findings of this study, the use of a calculator in mathematics classes should improve students' ability to learn mathematical concepts and apply calculator technology in an appropriate manner when solving mathematical problems.

(137 pages)

ACKNOWLEDGMENTS

There are several people who made this research possible whom I would like to thank. First is the American Educational Research Association/National Science Foundation *Evaluation Training Program* for providing the fellowship that made this work possible.

Next I would like to thank my committee—Dr. George Julnes, Dr. James S. Cangelosi, Dr. James T. Dorward, Dr. Steve M. Lehman, and Dr. Maria C. Norton—for all their insights and assistance. If I have any claim to being an effective teacher, I owe most of it to Jim Cangelosi. Special thanks to Dr. John C. Dossey, committee member at-large, who suggested that I research this issue. His counsel, patience, and encouragement kept me going when things got tough.

I would also like to thank the USU Psychology Department office staff—Sheila Jessie, Cathryn Peterson, Katie Hatch, and Karen Ranson—for working long-distance to make sure all my forms were properly completed, my "i"s dotted, "t"s crossed, and formatting was correct.

I owe a tremendous debt of gratitude to David Freund at Educational Testing Service. Without his extensive knowledge of the little details, nuances, and intricacies of NAEP data, it would have been impossible to carry out this work.

Finally, to my family—Amy, Dax, Claire, and Cy—whose patience, understanding, and support were more than should reasonably be asked of a family.

This work is dedicated to my foremothers and forefathers.

Kenneth L. Wareham

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF DEFINITIONS

Calculator active: Item requires calculator use; a student would likely find it almost impossible to solve the question without the aid of a calculator. See also *Calculator Appropriateness.*

Calculator allowed: NAEP mathematics content items/blocks which allowed students to use a calculator when solving if they so chose. For these items students were asked to indicate in their test booklets whether they did or did not use the calculator for each item.

Calculator appropriateness: Within the calculator allowed items each item is classified by its appropriateness for using a calculator to solve. Each item is designated within one of three classifications: *calculator active, calculator inactive,* or *calculator neutral.*

Calculator inactive: Items whose solution neither requires nor suggests the use of a calculator; in fact, a calculator would be virtually useless as an aid to solving the problem. See also *Calculator Appropriateness.*

Calculator neutral: Items in which the solution to the question does not require the use of a calculator, but some students might choose to do so. See also *Calculator Appropriateness.*

Calculator restricted: NAEP mathematics content items/blocks that did *not* allow students to use a calculator when solving.

Calculator use category: See f*requency of calculator use.*

Frequency of calculator use: Subject's reported use of a calculator from NAEP background question *M812711 How often do you use a calculator for math?* Response categories are *Almost Every Day* (*Daily*), *Once or Twice a Week* (*weekly*), *Once or Twice a Month* (*monthly*), and *Never or Hardly Ever* (*never*).

NAEP: National Assessment of Educational Progress.

NCES: National Center for Education Statistics.

NCTM: National Council of Teachers of Mathematics

CHAPTER I

INTRODUCTION

The renowned mathematician Gottfried Wilhelm von Leibniz once stated, "It is unworthy of excellent men to lose hours like slaves in the labor of calculation which could safely be relegated to anyone else if machines were used" (Goldstine, 1972, p. 8). The possibility of using a machine to automate mathematical calculations, eliminate toil, and create more time for leisure is certainly appealing. The development of such a device began as early as 500 B.C. with the counting board and continued to develop in such forms as the abacus, Napier's Rods, the Step Reckoner, and the slide rule. Perhaps the most significant event in the development of computational technology occurred in 1967 with the invention of the first electronic hand-held calculator. The hand-held calculator made it possible to perform mathematical calculations quickly, accurately, and with relative ease.

With the development of this technology came the question of the use and role of the calculator in schools and mathematics instruction. Some considered the calculator a powerful tool that could promote and increase learning; others considered it a crutch that would both create and support the mathematically feeble. For the past 35 years this protracted debate has been waged in research journals, professional societies, education policy meetings, newspapers, popular magazines, faculty meetings, and commercial testing companies.

Recognizing the potential of calculators, in 1974 the National Council of Teachers of Mathematics (NCTM) issued the following statement:

> With the decrease in the cost of the minicalculator, its accessibility to students at all levels is increasing rapidly. Mathematics teachers should recognize the potential contribution of the calculator as a valuable instructional aid. In the classroom, the minicalculator should be used in imaginative ways to reinforce learning and to motivate the learner as he becomes proficient in mathematics. (p. 3)

A year later the Conference Board of the Mathematical Sciences and National Advisory Committee on Mathematical Education (1975), supporting the position of the NCTM, suggested the following:

> Beginning no later than the end of the eighth grade, a calculator should be available for each mathematics student during each mathematics class. Each student should be permitted to use the calculator during all of his or her mathematical work including tests. (p. 138)

Advocates of calculators in educational settings have given the following reasons for their use (Hembree, 1984; Smith, 1996; Suydam, 1976):

1. They aid in evaluating, understanding, and learning algorithmic processes.

2. They facilitate concept development.

3. They enlarge the scope of problem solving by allowing realistic problems.

4. They greatly benefit student achievement in problem solving, especially for low-ability and high-ability students.

5. They motivate by encouraging discovery, exploration, and creativity.

6. Developing computational skill is not the central purpose of problem solving.

Notwithstanding the calculator's obvious advantages of speed and accuracy there are opponents to and arguments against its use, especially in early grades. The following reasons are often cited for *not* using calculators in educational settings (Klein, 2001; Pomerantz, 1997; Saxon, 1986; Suydam, 1976):

1. They are not available to all students.

2. They could be used as substitutes for paper and pencil skills.

3. They encourage the false impression that mathematics is largely mechanical and involves nothing more than computation.

4. Not enough research exists on their effects.

5. Basic mathematical skills decline if the calculator is used for computational purposes.

6. Students become dependent upon calculators, thus causing a further decline of mathematics in America.

7. Accurate assessment of student's skills cannot occur if calculators are permitted on tests.

Hembree (1984) addressed the first four arguments against using calculators and determined that (a) the argument on availability had dissipated when prices decreased, (b) items two and three were primarily straw-man arguments as few educators promoted calculator use at the expense of basic skills or knowledge of the nature of mathematics, and (c) argument four had merit and produced volumes of reports concerning the use and effects of calculators. The majority of these reports were primarily concerned with determining if calculators were detrimental to basic skills. The answer was usually "no," provided that students had learned the fundamentals using paper and pencil (Suydam, 1979, p. 3). Hembree then proceeded to address argument four using a relatively new procedure called meta-analysis (Glass, 1978). Hembree's integration of findings from 79 calculator studies found that using calculators had a positive effect on students' mathematical skills and attitudes, and did not produce adverse effects on students' paper-and-pencil skills. While generally applicable, this conclusion did not hold true at the fourth-grade level, where "basic skills significantly suffered" (Hembree, p. 173). Hembree's research also suggested that arguments five and six were uncertain and open to discussion.

Considering the volume of research supporting the use of calculators in mathematics classes, one might think that calculator use soon became the norm, but such was not the case. The NCTM's recommendation on calculator use met with stiff resistance and was generally not implemented, particularly in primary grades (Suydam, 1982). In addition, those opposed to calculator use found a forum for their emotive arguments in popular magazines and newspaper editorials where the general public would be more likely to read them, and in which there was little if any mention of what the research publications had to say on the issue. Today the calculator has become fairly commonplace in schools, but the debate over its use and effects continue to persist (National Council of Teachers of Mathematics, 1999).

Statement of the Problem

With the development of the electronic calculator came the debate of whether or not calculators should be used in mathematics education. Several experimental and quasi-experimental studies found conflicting results; some indicated that calculators had a positive impact on math achievement, others found no effect either way, and a few found negative effects.

As more research became available, state-of-the-art reviews and meta-analytical studies found a trend in overall results that supported calculator use. Though policies advocating calculator use initially encountered resistance, in time such policies were adopted and became the norm in mathematics classrooms.

With calculators now being commonplace in schools, the question shifts from *should* calculators be used to *what effect* has the implementation of calculator use policies had on students' mathematical achievement? Do the results of earlier studies from a particular place and time generalize to students across the nation today, or are the predictions of calculator dependency and decreased mathematical competence becoming a reality? Has the calculator actually become the technology that allows students to "learn more mathematics more deeply" (National Council of Teachers of Mathematics, 2000, p. 25), or have calculators fulfilled the ominous prediction of being a "crutch" to support those who have achieved "calculator-assisted mathematical incompetence" (Escobales & Rothenberg, 1987, p. 73).

A valuable source of information for assessing the results of large scale calculator implementation is the National Assessment of Educational Progress (NAEP). Commonly known as "The Nation's Report Card," NAEP assessments began in 1969 as a way to measure what America's students know and can do in various subject areas on a national level. The NAEP is conducted every 4 years in Grades 4, 8, and 12 and is recognized as "the largest and most comprehensive assessment of U.S. education, relating student achievement to instructional practices, teachers, principals, and school characteristics" (National Center for Education Statistics, n.d.). The 1996 NAEP

Mathematics Assessment included several items that could be used specifically to assess the effects of policies advocating the widespread use of calculators on mathematics achievement.

Purpose of the Study

The purpose of this study was to analyze the 1996 NAEP data to identify the relationship between calculator use and student performance on the NAEP Main Mathematics Assessment. This general purpose includes several sub issues. In addition to being interested in the overall relationship between use and NAEP achievement (including the effort to control for spurious factors), this study examined the contextual factors that moderate the impact of calculator use. Similarly, it analyzed the relationship between calculator use and student performance on calculator-allowed and calculator-restricted items, as well as the ability of students to recognize whether the use of a calculator was appropriate when responding to a math problem.

Research Questions

Four research questions were developed to assess the effects of policies advocating the widespread use of calculators in school classrooms.

Question 1: How does frequency of calculator use in the classroom relate to mathematics achievement on the NAEP Mathematics Assessment?

Question 2: How is this relationship with achievement affected when potentially confounding variables are controlled (e.g., teacher's education and experience, student's socioeconomic status, parent's education level, etc.)?

Question 3: How is this relationship affected when the data are disaggregated by question type, where the calculator is allowed on some NAEP questions but not others?

Question 4: How does frequency of calculator use relate to whether students recognize that it is appropriate or inappropriate to use a calculator to solve specific NAEP problems?

Significance of the Study

This study examines on a large scale the effects of policies implementing the widespread use of calculators in mathematics classrooms based on the results of prior calculator research. The results will be of value to educators and policymakers in determining the consequences of such policies. Consequences, as defined by Rogers (1995), are "the changes that occur to an individual or to a social system as a result of the adoption or rejection of an innovation" (p. 405). The consequences in this case are those predicted by the pro and con arguments within the calculator debate (note that consequences may be positive or negative).

Methodology

This research utilized the methods of quantitative analysis within a causal-comparative (also known as *ex post facto*) design to analyze the data from the 1996 NAEP Main Mathematics Assessment in order to determine the effect of policies advocating calculator use in schools on mathematics achievement.

The causal-comparative design was selected due to the fact that (a) the NAEP assessments do not involve experimental manipulation, (b) NAEP data is limited to secondary analysis procedures, and (c) the treatment is a categorical variable, not a continuous variable. These three factors are inherent limitations of NAEP data, but the advantage of using NAEP data is that it sets the assessment benchmark for relating student achievement to instructional practices, school characteristics, and education policies.

The assessment was compiled together in booklet form and contained between 30 to 45 items depending on which booklet the student received. The mathematics content items underwent an extensive writing, review, and field trial process prior to being used. The assessment contained a range of constructed-response and multiple-choice questions measuring performance on sets of objectives outlined by the National Assessment Governing Board (NAGB).

Assessments were conducted by Westat, Inc. using regional in-field assessment staff. The staff members conducted the assessment using standardized procedures to insure consistency and uniformity of administration. Assessments were administered between January 3 and March 29, 1996.

NAEP results were reported in scale scores and Research Questions 1 and 2 were analyzed using these scores. Research Questions 3 and 4 required items to be isolated by item type, and thus negated the conditions required for computing scaled scores. Due to this condition, Research Questions 3 and 4 were conducted using raw data to determine the percentage of items answered correctly. Summary statistics for each research question were reported along with the statistical significance of differences in scores and the practical significance (i.e., effect sizes) of score differences where applicable.

Delimitations

NAEP data has the benefit of being the largest and most representative sample of student achievement in the nation. According to the National Research Council (1996), NAEP data are

> an unparalleled source of information about the academic proficiency of U.S. students, providing among the best available trend data on the academic achievement of elementary, middle, and secondary students in core subject areas. In addition, NAEP has distinguished itself in setting an innovative and rigorous agenda for conventional and performance-based testing. (p. 5)

There are certain boundaries and limitations to the study that must be carefully considered when working with and interpreting NAEP results. These are summarized below:

1. One of the most important things to keep in mind is that:

> NAEP does not provide scores for individual students or schools; instead, it offers results regarding subject-matter achievement, instructional experiences, and school environment for populations of students (e.g., fourth-graders) and

subgroups of those populations, e.g., female students, Hispanic students (National Center for Education Statistics, n.d.).

2. The subjects selected for this study are from the eighth-grade level. Though the results may be applicable to eighth-grade students across the nation, they may not have the same generalizability to lower (e.g., K-6) or higher (e.g., 10-12) grades and should not be interpreted as having such widespread application.

3. NAEP contains two special characteristics that affect the validity of conventional techniques of statistical analysis. Specifically, the sampling procedures for selecting students are not simple random samples, but stratified multistage probability samples in which clusters of students are selected and certain subpopulations are sampled at a higher rate. Secondly, the use of a balanced incomplete block (BIB) spiraling of assessment items means that each examinee takes only a subset of the test items in any content area. These factors require special procedures for running computations that cannot be ignored. These factors are further outlined in the methods section and thorough descriptions are available in the NAEP technical manuals.

4. The *NAEP Data Tool* only permits factors to be examined one-at-a-time. Though it was often desirable to account for multiple factors at the same time, such analysis was beyond the capability of the current software.

5. NAEP assessments are subject to numerous variables that simply cannot be controlled. For the study at hand, several competing hypotheses such as student motivation, natural mathematical ability, effectiveness of instruction, availability of

resources, and so forth cannot be ignored or dismissed without cause. In addition, the internal validity threats of history and selection must be considered.

Recognizing and working within these conditions, NAEP data may be disaggregated by relevant factors that can be controlled in order to determine their relative effects on achievement. Operating under the null hypothesis that calculators have no effect on math achievement, there should be no differences in scores based on the calculator-use group, even when groups are further disaggregated by various relevant factors. If differences are found, it can provide meaningful insight as to "which subgroups are not responding in the way that others are—enabling us to understand why and to search for new processes so all students can learn" (Bernhardt, 2003, p. 36). For this study frequency of calculator use was held constant while the factors of parent's level of education, gender, socioeconomic status (SES), type of school, and NAEP achievement level were disaggregated. Though it was desirous to disaggregate data based on other variables such as time spent on homework and math course currently taking, limitations within the *NAEP Data Tool* did not allow for such analyses. Currently a new analysis tool with additional functionality, including regression and multiway cross tabulations, is being developed and is scheduled for release in the spring of 2005 (D. Freund, personal communication, December 14, 2004).

Finally, it should be noted that this study was started after the 2000 NAEP assessment was administered but before the 2000 NAEP data was available; as a result, this study uses the data from the 1996 assessment. Now that this study is near completion, the 2000 data is available but not feasible for acquisition, analysis, and inclusion in this work. Where available and pertinent, the results of the 1996 and 2000 administrations were compared for differential affects between years. These analyses were limited to data available in the *NAEP Data Tool*, therefore results from Research Questions 3 and 4 could not be compared between 1996 and 2000 datasets. In nearly all situations the between-year comparisons were consistent with each other and any differences were not statistically significant.

Organization of Remaining Chapters

With the conclusion of this general overview, the remainder of the study is presented in four chapters. Chapter II is a review of literature related to the use of calculators in schools and their impact on mathematics achievement. Chapter III is a detailed description of the methods used in conducting this study. Chapter IV presents the findings and overview of results. Chapter V concludes the work with a summary and discussion of the results and their implications in the field of mathematics education.

CHAPTER II

REVIEW OF RELATED LITERATURE

A significant body of research concerning the calculator's effects on mathematics achievement provides the basis for this study. This chapter will begin with an overview of the history and development of the calculator, explain the research process in reviewing the literature, and examine the empirical studies in the field.

History and Development of Calculators

Efforts to produce a mathematical calculating machine can be accurately dated back to at least 300 B.C. The first known devices were called counting boards, the oldest surviving example being the Salamis tablet discovered in 1846 on the Island of Salamis (Fernandes, 2001). Counting boards evolved into what we know today as the Chinese abacus. The abacus dates back to 1200 A.D. and is believed to have been brought to the east by early Christians (Fernandes, 2001). During the Middle Ages the abacus was replaced by arithmetic (counting using written numbers) throughout most of Europe. In 1617 John Napier invented a calculating machine known as Napier's rods, or Napier's bones. Napier's rods were quite popular in their day throughout Europe and continued to be used in British schools up until the 1960s (Diploudis, 1997). Though these devices were useful, they were not necessarily calculating machines because they simply indicated results that were actually worked out in the mind of the operator.

The first mechanical device that truly calculated a result was invented in 1641 by the French mathematician Blaise Pascal (Michaelson, 1997a). Commonly known as a "Pascaline," this complex machine could sum up to six-digit numbers. The Pascaline could readily sum figures, but subtraction, multiplication, and division were complex and limited operations. The excessive price, difficulty of operation, and propensity for mechanical failure limited the sale and use of the Pascaline so that it never really became popular (Michaelson, 1997b). In 1673 Leibniz used the Pascaline as the basis for his own computation machine, which readily produced results using addition, subtraction, multiplication, or division without the complex operations required by the Pascaline. The basic concept of Pascal's machine can still be found in contemporary mechanical adding machines.

A major advancement in calculator technology came in the form of the Difference Engine. A small hand-cranked machine built by Charles Babbage in 1822, the Difference Engine was capable of generating logarithmic and astronomical tables to an accuracy of six decimal places. The difference engine operated using punch card programming, had a memory of one thousand 50-digit numbers, and produced visual readouts. Babbage died before the Difference Engine went into production, but his contributions provided the basis and foundation for the development of modern calculating instruments (Moursand, 1981; Science Museum, 2001).

The first electronic calculator, the Electronic Numerical Integrator and Calculator (ENIAC), was developed in 1946 at the University of Pennsylvania by Presper Eckert and John Mauchly. (It could be argued that the first electronic calculator was the Colossus machine built in 1943 and used at Bletchley Park to crack the German Enigma and Lorenz codes, but due to the top-secret nature of Colossus, the ENIAC is generally recognized as the first electronic calculator.) The ENIAC contained 19,000 vacuum tubes, 1,500 relays, hundreds of thousands of resistors, capacitors, and inductors, consumed almost 200 kilowatts of electrical power, and was capable of multiplying two ten-digit numbers in 2.6 milliseconds (Weik, 1961).

Developments in technology allowed vacuum tubes and transistors to be replaced with silicone. This new technology led to the mass production of electronic calculators, which began in 1965 (Moursand, 1981). These machines contained $170 worth of electronic components, were hand assembled, and sold for $1,500. In 1967 Jack S. Kilby, Jerry Merryman, and Jim Van Tassel, working at Texas Instruments Inc., invented the first electronic hand-held calculator. The technology now allowed for a reasonably affordable device that could compute calculations both quickly and accurately (Moursand). Due to their small size and relative affordability, calculators were becoming commonplace by the early 1970s (Williams, 1978). Today's calculators have tremendous computing capability, especially the scientific, programmable, and graphing models.

Conducting the Review of Research

The literature review started with an on-line computer search of the following databases: Educational Resources Information Clearinghouse (ERIC), PsychINFO, Current Index to Journals in Education (CIJE), and Dissertation Abstracts International

(DAI). Using the subject terms *mathematics* and *education,* along with the key word *calculator*, a list of references was produced. This list was pared down using the key words *meta-analysis*, and *review of research* in order to find reviews of the major research studies concerning the use and effects of calculators in mathematics education. This list included two meta-analyses (Hembree, 1984; Smith, 1996) and a series of state of the art reviews by Suydam from the Calculator Information Center at Ohio State University (Suydam, 1978, 1979, 1980, 1981, 1982). A reading of these sources led to a more focused topic and more specific literature search dealing with calculator use at the middle school level and its effects on achievement (e.g., use of the key words *middle school, junior high, 6^{th}-9^{th} grade, achievement, scores,*). The final body of studies included journal articles, ERIC reports, unpublished reports, conference proceedings, dissertations, newspaper reports, and periodicals. Studies that were available in electronic format were downloaded, the rest were obtained in hardcopy format from the library.

Review of Calculator Studies

As advances in calculator technology took place, making them both more available and affordable, the number of studies to examine their effect on learning increased. One of the first studies attempting to determine the relationship of calculators to mathematical achievement was conducted by Emmett Betts in 1937 (Shult, 1987). Betts hypothesized that students would be more accurate and efficient problem solvers if they used calculators in mathematics. To test this theory he selected 13 above average 6th grade students to participate in a 6-week treatment program.

Betts administered the students a pretest in order to establish a baseline for comparison before starting treatment. The pretest consisted of an experimenter-designed instrument that included operational problems of whole numbers, fractions, and decimal numbers. During the 6-week study, students were allowed to use calculators at will. At the end of the treatment a posttest similar to the pretest was administered and scores of the two tests compared. Betts discovered that all students scored higher on the posttest than the pretest (Betts, 1937). Because no control or comparison group was included in this study, the students selected were considered above average, and other history threats to internal validity, Betts was prevented from drawing any authoritative conclusions from his work.

Fehr, McMeem, and Dobel (1957) conducted a pretest-posttest, control group design using fifth-grade students to study calculator effects on paper-and-pencil computation and mathematical reasoning. Both groups received instruction in the same content for 4 months, with the treatment group being able to use calculators to supplement the material. Their conclusion was that the experimental group gained more than the control group with respect to both reasoning and computation skills (Fehr et al.).

Using a similar experimental design, Durrance (1964) studied the effect of calculators on mathematics achievement using 70 sixth-, seventh-, and eighth-grade students matched on IQ and math achievement. Students were randomly assigned to treatment and control groups, then given a pretest on mathematical computations without using calculators. Over the next 3 weeks the two groups were given the same instruction, with the experimental group being allowed to use calculators. Following the math unit a posttest was administered. Analysis of the scores found no statistically significant differences between the two groups in terms of math achievement.

In the mid-1970s research on calculators took center stage and produced "one of the largest bodies of research on any topic or material in mathematics education" (Suydam, 1982, p. 1). As calculator research became increasingly popular, Suydam became the leading chronicler of the calculator research field. Her initial work provided to the National Science Foundation a status report of calculators in precollege education (Suydam, 1976). Suydam used literature searches and questionnaire surveys to compile arguments on the pros and cons of calculator usage. Suydam's work revealed information on calculator usage, parent/teacher attitudes towards their use in the classroom, and research on student effects. These reports provided modest amounts of useful information in terms of research due to the fact that most of the investigators described their work as preliminary studies, inquiry, or exploration. In addition, short treatment periods, small sample sizes, and threats to internal validity limited the ability to generalize results to the population at large. The reports did seem to indicate that calculators could be used to teach certain topics, but it was not clear that such methods would result in achievement gains (Suydam, 1976).

From 1978 to 1982 the Calculator Information Center at Ohio State University issued a state-of-the-art review, which was authored by Suydam. The first four reports focused on progress toward acquiring and implementing calculators in schools, along with the ways in which the devices were being used (Suydam, 1978, 1979, 1980, 1981). These reports document that the availability of calculators to precollege students was increasing, while at the same time resistance to their use decreased. At elementary levels the primary usage was for a) drill of basic facts, b) checking paper-and-pencil answers (this was the most frequently cited use), c) games, d) direct calculation in problems, and e) exploring mathematical ideas. In secondary schools the emphasis was on a) direct calculation in problems, b) games, c) exploration, and d) use of textbook. The majority of these studies were aimed at determining if calculators were detrimental to the acquisition and retention of basic skills. Suydam's reviews concluded that the answer was no, as long as the fundamentals were established using paper and pencil (Suydam, 1979, p. 3).

Suydam's final review was explicitly devoted to a summary of research (1982). To date some 150 documents had been collected on various topics in calculator research. This summary found that with respect to achievement measures, 43 studies showed higher scores for calculator groups, 47 found no difference, and 5 favored the noncalculator groups.

Suydam's work primarily chronicles the implementation of calculators in the classroom. Other reviews have been conducted that are more specific. Parkhurst (1979) reviewed 9 studies at the junior high level, most of which found no statistically significant differences in achievement. Where differences were found, the advantage favored the calculator groups. Though Parkhurst's review showed promise for using calculators, he also points out the potential for significant bias in the results due to variability in teachers and the technology novelty effect.

Roberts (1980) assessed 37 studies ranging from elementary school through college. For elementary grades, 6 of 11 studies favored the calculator groups, even though they were not allowed to use calculators on tests. Similar results occurred in 6 of 11 studies at the secondary level, and 7 of 8 at the college level. The remaining seven studies found no statistically significant differences between groups. Roberts' cautioned against making conclusive judgments from his work, citing defective research designs, uncontrolled teacher variables, the absence of calculators on the posttest, and noncalculator students using calculators outside of class.

Rabe (1981) reviewed findings from 26 studies. The results showed that 14 of the studies favored calculator groups, 10 found no difference, and 2 found greater achievement for noncalculator groups.

Neubauer (1982) looked at seven studies and concluded that using calculators prior to junior high was ill advised. His findings indicated that students need to understand "the basics" before using calculators. He made the same recommendation for low-achieving students.

Sigg (1982) evaluated 22 studies and found that achievement scores from calculator groups were equal to or better than scores from their noncalculator comparison groups.

The reviews of research listed to this point are of the narrative type, with occasional studies using vote-counting methods. These reviews are subject to the faults inherent in these types of research integration (i.e., quality of research design, disregarding of sample size, equivalent weighting of nonequivalent studies without regard to differences in the magnitude of effects, etc.). Noting the shortcomings in these early methodologies, Hembree (1984) set out to bridge the gaps in these reviews using a relatively new method of research synthesis known as meta-analysis (Glass, 1978). Hembree's work has since become a landmark study in calculator research.

Hembree (1984) located studies using computer searches in *ERIC* and *DAI* data bases; manual searches in *CIJE*, *DAI*, and *Journal of Research in Mathematics Education* from 1972 to 1984; and direct requests for references from the Calculator Information Center at Ohio State University.

From the titles found, Hembree limited selection to only those works that (a) used students in grades K-12, (b) utilized electronic hand-held or desktop calculators, (c) contained group means and standard deviations, (d) provided continuous outcome data, and (e) contained a sample size of at least 10 subjects. In addition, no study would be rejected on the grounds of a flawed design (1984, pp. 125-126). The end collection of reports contained 79 studies including 12 journal articles, 12 ERIC documents, 53 dissertations, one project report, and one unpublished report.

Hembree (1984) addressed 15 specific research questions, of which the following are pertinent to this study:

1. What are the calculator's effects regarding acquisition of composite operational and problem solving skills?

2. What are the calculators effects regarding retention of operational and problem solving skills?

In answer to these questions, Hembree concluded:

> Regarding composite operational skills, nonsignificant effects existed for low and high ability students, while general students (in regular classes) produced significant effects related to school grade level, i.e., -.152 for grade 4 and .137 for the other grades combined. Hence, paper-and-pencil skills of low and high ability students in the calculator groups remained at par with basic skills of corresponding students in the control groups. In mixed ability classes, paper-and-pencil skills significantly improved from calculator treatment, except in grade 4 where basic skills significantly suffered.
>
> Paper-and-pencil achievement of low and high ability students did not change as a result of calculator treatment, but basic skills improved in general classes (effect size = .124), except in grade 4 (descriptive effect = -.181).
>
> ... calculator usage yielded achievement as high or higher than when calculators were not used and concept acquisition from the use of calculators appeared to be minimal.
>
> Confirming expectation that a use of calculators on tests will improve student scores, effects for basic operand and composite problem solving were consistently large and positive across grade level...Low and high ability effects (.436 and .458) appeared significantly higher than the descriptive effect (.271) for general students.
>
> ... Analyses of results for productivity were not conclusive (though a trend perhaps existed toward solution of more problems by the calculator groups). The extension effect for selectivity was fairly large and positive (.328). Hence, the calculator's use in problem solving created not only a computational advantage but also a benefit (probably time) in choosing proper approaches to solutions. (Hembree, 1984, pp. 173-175)

Based on the results of this meta-analysis Hembree (1984, pp. 178-179) recommended that:

1. Calculators be used in all math classes from Grades K-12, with levels of usage increasing as grade level increases.

2. Due to limited research for Grades K-3 and to the apparent negative effect of calculator treatments in Grade 4, the use of calculator in those grades should be restricted to familiarization, recreation, and perhaps occasional drill and problem solving.

3. Students in Grade 5 and beyond should be permitted calculator use in all problem-solving activities, including testing situations.

4. Teachers should prepare themselves for calculator instruction through self-training and in-service programs.

5. Curriculum developers should determine how the calculator can be optimally absorbed within the existing curriculum, and where existing curriculum should be revised to accommodate optimal calculator usage.

Of course, not all researchers and educators agreed with Hembree's (1984) conclusions and recommendations. Over the next 12 years 30 additional studies were conducted in an attempt to provide educators with conclusive indication of the best use of calculators in the development of mathematical skills (Smith, 1996).

Smith (1996) replicated Hembree's (1984) meta-analysis study using studies conducted since 1984. One significant difference between Smith and Hembree's work was the introduction of the graphing calculator. Now technology could not only compute algorithms, but could also graphically display lines and curves as well as plot data points in the Cartesian plane.

Overall, Smith's conclusions and recommendations were similar to Hembree's (1984), but his results were slightly different. Smith (1996) found no statistically significant differences in overall achievement of students in Grades 4, 5, 6, and 11, and statistically significant differences favoring calculator users in Grades 3, 7, 8, 9, and 10 (compare to Hembree who found negative effects at Grade 4 and positive effects for all other grades).

With the amount of research that had been done on calculators one might think that calculators would make their way into the curriculum, but for several years during and after these research efforts many teachers refused to use calculators in the mathematics instruction. A 1982 report by Suydam indicated that less than 20% of elementary and 36% of secondary teachers employed calculators in the classroom (Suydam, 1982, p. 3).

The debate over calculators continues to be waged. In *The Continuing Calculator Controversy,* Thomas Dick (1988) proclaimed that the argument concerning calculator use has more to do with "image than substance." (p. 37) Dick wisely pointed out that the opponents of calculators see their predicted consequences (i.e., students blindly punching buttons without using estimation or number sense to judge the results) as inevitable. Such conduct, claim calculator opponents, will inhibit the learning of basic skills needed to perform everyday mathematical problems and impede students' learning of more advanced mathematics. Dick empathized with this point, claiming "… the image of the calculator being used indiscriminately in the classroom, with no purpose other than to furnish students with a 'black box' with which to perform arithmetic calculations, should be objectionable to any responsible educator" (p. 38). He then quotes from Suydam (1976), who stated that "few educators believe that children should use calculators in place of learning basic mathematical skills" (p. 38), and finished off by pointing out that "organizations like the NCTM have *never* [author's emphasis] advocated using calculators as a substitute for instruction in estimation and basic arithmetic skills" (p. 39).

When considering the arguments of those who oppose the use of calculators some interesting facts emerge. First, many of the articles' opponents cite supporting their position are the same articles used by those who support calculator usage; they just have a different interpretation of the results, or only cite sections of the study that support their position (Dick, 1988; Saxon, 1986).

Second, many of the arguments are based on personal experience and testimonials rather than research (Gelernter, 1998; Hunsaker, 1997). It may be easy to recall an example of a mistake in the check-out line or people indiscriminately reaching for a calculator to do an "easy" computation, but such examples, spurred-on by the base-rate fallacy, serve to overgeneralize the belief that calculators have reduced people to mathematical incompetence. According to Brehm, Kassin, and Fein, "As long as a personal anecdote is seen as relevant, and the source as credible, it seems that one good image is worth a thousand numbers" (1999, p. 105). (Note that in the references cited above Gelernter is a professor of computer science at Yale University and Hunsaker is a math tutor and adult education teacher in Santa Clara, California.)

Finally, the articles against calculators are seldom written in scholarly journals, but rather in popular magazines where they reach a much larger audience, and one that is generally unfamiliar with systematic, scientific research. The arguments in these writings generally make an appeal to "common sense," emphasize one of several fallacies such as "calculator as crutch," "calculators think for the student," and that it "causes calculator dependency" (Pomerantz, 1997), or make the claim of "What was good enough for me should be good enough for (kids today)" (Dick, 1988, p. 39).

The following reasons are often cited for not using calculators in educational settings (Klein, 2001; Pomerantz, 1997; Saxon, 1986; Suydam, 1976):

1. They are not available to all students.

2. They could be used as substitutes for paper-and-pencil skills.

3. They encourage the false impression that mathematics is largely mechanical and involves nothing more than computation.

4. Not enough research exists on their effects.

5. Basic mathematical skills decline if the calculator is used for computational purposes.

6. Students become dependent upon calculators, thus causing a further decline of mathematics in America.

7. Accurate assessment of student's skills cannot occur if calculators are permitted on tests.

The first argument would eventually be proven null and void as the cost of calculators became trivial and most schools provided one for students who could not afford their own (Bracey, 1998). Arguments two and three are generally considered moot points as few knowledgeable educators promote calculator use at the expense of basic skills or knowledge of the nature of mathematics (Bracey; Dick, 1988; National

Council of Teachers of Mathematics, 2000). The fourth argument, at the time, was certainly valid, and produced a large number of studies to determine the impacts of using calculators in Grades K-12. The research consistently seemed to support calculator usage, provided that the fundamentals had first been established using paper and pencil (Suydam, 1979). The last three arguments remain points of contention to this day. The May/June 1999 issue of *Mathematics Education Dialogues* was dedicated to calculator usage and is appropriately titled *Groping and Hoping for a Consensus on Calculator Use* (National Council of Teachers of Mathematics, 1999). Once again, the supporters and detractors squared off in this continuing debate, with neither side emerging as the clear winner.

Summary of Reviews

After more than 30 years of investigation and some 200 studies, the majority of research supports the notion that students can learn more mathematics more deeply with the appropriate use of technology. General consensus is that calculators, at worst, have no adverse effect on student achievement provided students understand arithmetic, have a firm grasp of basic skills, and are able to assess the reasonableness of the calculator's computations.

Arguments that oppose the use of calculators are primarily based on anecdotal evidence, personal experience, and other nonresearch based opinions. Despite these primarily rhetorical and emotive types of arguments, and research to the contrary, there continues to exist a perception that calculators are detrimental to students' learning and achievement in mathematics.

A limitation within the prior research is the generalizability of the research findings. Though the large amount of research supporting calculators would seem to support generalizability, it does have some limitations. First of all, the studies lasted anywhere from one day to one year, and secondly, the median sample size was 30 students.

One untapped source of data to check the generalizability of the calculator research is NAEP. The 1996 NAEP was administered to over 7,000 students from across the nation. Though it cannot be determined exactly how long students who took the NAEP had been using a calculator, the timing of the assessment (February-March, 1996) implies that those who reported using a calculator would have been doing so since at least the beginning of the 1995-1996 school year (5 to 6 months minimum). Finally, the 1996 NAEP Mathematics Assessment included several items that could be used specifically to assess the effects of policies advocating the widespread use of calculators on mathematics achievement. The availability of these data and the desire to assess the effects of calculator use on a national level provided the impetus for this study.

CHAPTER III

METHODOLOGY

This chapter explains the methods used in carrying out the study. The reader is reminded of the unique characteristics and considerations for working with NAEP data mentioned in Chapter I. These features will receive further elaboration in this section.

Limited NAEP data is available online and may be readily accessed using the *NAEP Data Tool.* The full NAEP data set is only available to qualified institutions and requires a *Restricted Use Data License* available from NCES. For researchers working with restricted use data, NCES offers a 4-day NAEP training session. This workshop covers such items as NAEP history, item and instrument development, data collection procedures, technical issues associated with BIB spiraling and sample weighting, and the use of WESVAR, NAEPEX, and NAEPREG software for selecting and extracting data. This author attended the training session in July of 2000.

When this study was undertaken, the 2000 NAEP had been administered but it would take considerable time before the booklets were processed and the data made available. Due to this situation, the research herein was conducted using data from the 1996 NAEP. As this study was being completed, the 2000 NAEP data was released but it was neither practical nor feasible to acquire, analyze, and include in this work. In certain situations it was prudent to perform some analyses using the 2000 data to determine if any significant changes had occurred between the two administrations. Such situations are described where applicable.

For reference purposes the research questions are restated here:

Question 1: How does frequency of calculator use in the classroom relate to mathematics achievement on the NAEP mathematics assessment?

Question 2: How is this relationship with achievement affected when potentially confounding variables are controlled (e.g., teacher's education and experience, student's socioeconomic status, parent's education level, etc.)?

Question 3: How is this relationship affected when the data are disaggregated by question-type, where the calculator is allowed on some NAEP questions but not others?

Question 4: How does frequency of calculator use relate to whether students recognize when it is appropriate or inappropriate to use a calculator to solve specific NAEP problems?

The General Perspective

This research utilizes quantitative methods within the causal-comparative (aka ex post facto) design in order to assess the cause-effect relationship between calculator use and achievement in mathematics. The causal-comparative design was selected because (a) the NAEP assessments do not involve experimental manipulation, (b) NAEP data is limited to secondary analysis procedures, and (c) the treatment is an ordinal variable. These conditions, as well as the fact that the causes are being studied after they have had their presumed effect, make this study well suited for using the causal-comparative design (Gall, Borg, & Gall, 1996).

The Research Participants

The subjects for this research are taken from the 1996 NAEP Main Mathematics Exam, eighth grade level, with no accommodations permitted. This sample was chosen for reasons associated with content items, sample size requirements, and the nature of providing accommodations during assessments.

Prior to 1996 calculators were only a minor aspect of NAEP, but increased use forced the NAGB to consider calculators as an issue that warranted increased consideration (Mullis, Dossey, Owen, & Phillips, 1991; National Assessment Governing Board, 1994). For the 1996 assessment, approximately one third of the items permitted the use of a calculator and the subject-specific background questions included items specifically designed to assess the use of calculators by both students and teachers (National Assessment Governing Board, 2002).

Because of minimum *N* size requirements, it was important to select a sample that would have a high probability of meeting the minimum sample size in each calculator use category. According to the *NAEP 1996 Technical Report* (National Center for Education Statistics, 1999, p. 193):

> For results to be reported for any subgroup, a minimum sample size of 62 was required. This number was arrived at by determining the sample size required to detect an effect size of 0.5 with a probability of .8 or greater. The effect size of 0.5 pertains to the "true" difference in mean proficiency between the subgroup in question and the total population, divided by the standard deviation of proficiency in the total population. In addition, subgroup members must represent at least five Primary Sampling Units.

An exploratory analysis of the data was used to determine the number of students in each calculator use subcategory. The results of this analysis are shown in Table 1.

The data in Table 1 indicates that the 4^{th}-and 12^{th}-grade samples were at the greatest risk for not meeting the minimum sample size criteria; 4^{th} grade due to its low percentage in the *daily* subgroup (10%), and 12^{th} grade in both the *monthly* (7%) and *never* (9%) subgroups. The eighth grade sample has two factors that make it appealing; the largest *N* size (7033), and the highest percentage of students in its lowest subgroup (12% in *never*). These factors give it the highest probability of having enough students in each subgroup to allow for reliable estimates and valid interpretations.

In 1996 NAEP began to provide the inclusion/accommodation criteria for students with learning disabilities (SD) or limited English proficiency (LEP) that was typically provided by their school. As this was a new and somewhat experimental feature, NAEP officials divided the school sample into three subsamples in order to

Table 1

Student Reported Frequency of Calculator Use by Grade Level.

		Distribution of students who reported using a calculator within each of the following frequency categories			
Grade	*N*	*Almost every day*	*Once or twice a week*	*Once or twice a month*	*Never or hardly ever*
4	6,523	10 %	23 %	26 %	41 %
8	7,033	48 %	26 %	14 %	12 %
12	6,832	69 %	15 %	7 %	9 %

determine the effects of the new provisions. These subsamples were defined as follows (National Center for Education Statistics, 1999, p. 105):

Sample 1. These schools used the inclusion/accommodation criteria from 1990 and 1992, and accommodations were not provided.

Sample 2. These schools used the new 1996 inclusion/accommodation criteria, but accommodations were not offered.

Sample 3. These schools applied the new 1996 criteria and the accommodations most commonly used for achievement testing were offered.

The research presented herein was conducted using subsample 2, which is identified by its designation as "1996^{n}" in the *NAEP Data Tool* and as Reporting Sample 1 in the NAEP Restricted Use Data.

NAEP Sampling Procedures

NAEP went through considerable efforts to insure that selected participants were representative of the nation's student population and subgroups of that population. The sampling design used a complex multistage process that relied on stratification to insure adequate representation. A brief description of the sampling procedure is provided below. For complete details of the sampling procedure see chapters 1, 3, and 5 of *The NAEP 1996 Technical Report* (National Center for Education Statistics, 1999).

The first step in selecting the sample was to divide the nation into primary sampling units (PSU). Each PSU is contained within one of four regional areas and designed to meet a minimum size requirement based on population. These regions were used to stratify the PSU, ensuring that each region was adequately represented in the assessment sample. The 22 largest PSUs were included in the sample with certainty due to their size and population characteristics. Seventy-two PSUs were selected from the rest of the nation using sample weighting methods that insured adequate representation of Black and Hispanic students.

In the second stage of sampling, public and private schools within selected PSUs were randomly selected for participation. Again, stratified sampling with weighting for accurate representation of Black and Hispanic students was used.

The third and final sampling stage required the generation of a list of all grade eligible students within the selected schools. Students from this list were randomly selected to participate in the assessment. Participation rates for the 8th grade main mathematics assessment were as follows: school participation, 81.5%; student participation, 92.9%; overall participation, 75.7% (National Center for Education Statistics, 1999, p. 72).

Despite NAEP's goal of assessing all selected students, certain students who were judged by school authorities as being incapable of meaningful participation in the assessment were excluded from the selection pool.

When the sampling process was completed 7,146 eighth-grade students were selected to take the main mathematics assessment. (Note: Of these students, 109 did not respond and 4 gave multiple responses to the background question on calculator use. These students were removed from the sample and the final N size was 7,033 as indicated in Table 1.)

The demographic characteristics of the selected students are reproduced from the NAEP Technical Report (National Center for Education Statistics, 1999, p. 387) and presented in Table 2.

Instruments/Tasks and Materials

The instrument used to measure mathematics achievement is the 1996 National Assessment of Educational Progress 8th Grade Main Mathematics Assessment. A general description of the assessment is given below; for detailed information see Chapter 4 of The *NAEP 1996 Technical Report* (National Center for Education Statistics, 1999). Emphasis is given to the fact that all of the following information on assessment items, instruments, and administration was designed, developed, conducted by NAEP; this author takes neither credit nor responsibility for their work.

The assessment is given in booklet form and contains general background questions, subject-specific background questions, and mathematics content items in multiple choice and constructed response formats. All items are in print form and completed with a No. 2 pencil.

The mathematics content items underwent an extensive writing, review, and field trial process. Following field trials, experts from state education agencies and Educational Testing Service analyzed the results. Based on these analyses items were revised, modified, or edited where necessary and subjected to a second review and field test. After a final review by the Instrument Development Committee to ensure that the items had fully met all criteria they were printed and bound into booklets.

Table 2

Demographic Characteristics of Students Selected for the 1996 NAEP Eighth-Grade Main Mathematics Assessment

Demographic data	*N*	Percent
Gender		
Male	3,597	50.3
Female	3,549	49.7
Race		
White	4,501	63.0
Black	1,193	16.7
Hispanic	911	12.7
Asian American	408	5.7
American Indian	110	1.5
Unclassified	23	0.3
Region		
Northeast	1,312	18.4
Southeast	1,883	26.4
Central	1,726	24.2
West	2,225	31.1
Type of location		
Central city	3,218	45.0
Urban/large town	2,186	30.6
Rural/small town	1,742	24.4
School type		
Public	5,590	78.2
Nonpublic	1,556	21.8
Modal age		
Younger	48	0.7
At modal age	4,380	61.3
Older	2,718	38.0

The booklets contained three blocks of mathematics content items. Each block was designed to take 15 minutes to complete, thus blocks with constructed response questions were likely to have relatively few items, while blocks composed primarily of multiple choice questions would have a relatively higher number of items. The typical booklet contained between 30 and 45 mathematical content items in total.

The blocks were arranged in a BIB design with “spiraled” administration (National Center for Education Statistics, 1999, p. 75). The BIB spiraling process was used to maximize the possibility that all items had an equal chance of being presented and answered by the examinees (Deng, Ferris, & Hombo, 2003).

The items used in the mathematics content questions contained a range of constructed-response and multiple-choice questions measuring performance on sets of objectives outlined by NAGB. All mathematics items were classified using a three-by-five matrix of content strands and ability levels. The content strands are categorized and described as follows (National Center for Education Statistics, 1999, pp. 32-33):

> *Number Sense, Properties, and Operations*
>
> This strand focuses on students’ understanding of numbers (whole numbers, fractions, decimals, integers, real numbers, and complex numbers), operations, and estimation, and their application to real-world situations. Students will be expected to demonstrate an understanding of numerical relationships as expressed in ratios, proportions, and percents. Students also will be expected to understand properties of numbers and operations, generalize from number patterns, and verify results.
>
> *Measurement*
>
> The measurement strand focuses on understanding of the process of measurement and on the use of numbers and measures to describe and compare mathematical and real-world objects. Students will be asked to identify

attributes, select appropriate units and tools, apply measurement concepts, and communicate measurement-related ideas.

Geometry and Spatial Sense

As described in the NCTM *Standards*, spatial sense must be an integral component of the study and assessment of geometry. Understanding spatial relationships allows students to use the dynamic nature of geometry to connect mathematics to their world.

This content strand is designed to extend well beyond low-level identification of geometric shapes into transformations and combinations of those shapes. Informal constructions and demonstrations (including drawing representations), along with their justifications, take precedence over more traditional types of compass-and-straightedge constructions and proofs. While reasoning is addressed throughout all of the content strands, this strand continues to lend itself to the demonstration of reasoning within both formal and informal settings. The extension of proportional thinking to similar figures and indirect measurement is an important connection here.

Data Analysis, Statistics, and Probability

The important skills of collecting, organizing, reading, representing, and interpreting data will be assessed in a variety of contexts to reflect the pervasive use of these skills in dealing with information.

Statistics and statistical concepts extend these basic skills to include analyzing and communicating increasingly sophisticated interpretations of data. Dealing with uncertainty and making predictions about outcomes require an understanding not only of the meaning of basic probability concepts but also the application of those concepts in problem-solving and decision-making situations.

Questions will emphasize appropriate methods for gathering data, the visual exploration of data, a variety of ways of representing data, and the development and evaluation of arguments based on data analysis. Students will be expected to apply these ideas in increasingly sophisticated situations that require increasingly comprehensive analysis and decision making.

Algebra and Functions

This strand extends from work with simple patterns at grade 4, to basic algebra concepts at grade 8, to sophisticated analysis at grade 12, and involves not only algebra but also pre-calculus and some topics from discrete mathematics. As described in the NCTM *Standards*, these algebraic concepts are developed throughout the grades with informal modeling done at the elementary level and with increased emphasis on functions at the secondary

> level. The nature of the algebraic concepts and procedures included in the assessment at all levels reflects the NCTM *Standards*. Students will be expected to use algebraic notation and thinking in meaningful contexts to solve mathematical and real-world problems, specifically addressing an increasing understanding of the use of functions (including algebraic and geometric) as a representational tool.

NAEP ability levels (i.e., difficulty levels) were defined under the auspices of the NAGB an are described as follows (National Center for Education Statistics, 1999, p. 34):

> *Conceptual Understanding*
>
> Students demonstrate conceptual understanding in mathematics when they provide evidence that they can recognize, label, and generate examples and nonexamples of concepts; use and interrelate models, diagrams, manipulatives, and varied representations of concepts; identify and apply principles (i.e., valid statements generalizing relationships among concepts in conditional form); know and apply facts and definitions; compare, contrast, and integrate related concepts and principles to extend the nature of concepts and principles; recognize, interpret, and apply the signs, symbols, and terms used to represent concepts; or interpret the assumptions and relations involving concepts in mathematical settings.
>
> Conceptual understanding reflects a student's ability to reason in settings involving the careful application of concept definitions, relations, or representations of either. Such an ability is reflected by student performance that indicates the production of examples, common or unique representations, or communications indicating the ability to manipulate central ideas about the understanding of a concept in a variety of ways.
>
> *Procedural Knowledge*
>
> Students demonstrate procedural knowledge in mathematics when they select and apply appropriate procedures correctly; verify or justify the correctness of a procedure using concrete models or symbolic methods; or extend or modify procedures to deal with factors inherent in problem settings.
>
> Procedural knowledge includes the various numerical algorithms in mathematics that have been created as tools to meet specific needs efficiently. Procedural knowledge also encompasses the abilities to read and produce graphs and tables, execute geometric constructions, and perform non-computational skills such as rounding and ordering. These latter activities can be differentiated from conceptual understanding by the task context or presumed student

background—that is, an assumption that the student has the conceptual understanding of a representation and can apply it as a tool to create a product or to achieve a numerical result. In these settings, the assessment question is how well the student executed a procedure or how well the student selected the appropriate procedure to effect a given task.

Procedural knowledge is often reflected in a student's ability to connect an algorithmic process with a given problem situation, to employ that algorithm correctly, and to communicate the results of the algorithm in the context of the problem setting. Procedural understanding also encompasses a student's ability to reason through a situation, describing why a particular procedure will give the correct answer for a problem in the context described.

Problem Solving

In problem solving, students are required to use their accumulated knowledge of mathematics in new situations. Problem solving requires students to recognize and formulate problems; determine the sufficiency and consistency of data; use strategies, data, models, and relevant mathematics; generate, extend, and modify procedures; use reasoning (i.e., spatial, inductive, deductive, statistical, or proportional) in new settings; and judge the reasonableness and correctness of solutions. Problem solving situations require students to connect all of their mathematical knowledge of concepts, procedures, reasoning, and communication/ representational skills in confronting new situations. As such, these situations are, perhaps, the most accurate measures of students' proficiency in mathematics.

The items in each booklet were selected from a pool of 162 items of the following types and amounts: multiple-choice, 91; constructed-response dichotomously scored, 26; constructed-response polytomously scored, 42; cluster items, 3 (National Center for Education Statistics, 1999). It should be noted that cluster items are a series of objective questions based on a common stem, thus one cluster item could contain as many as six dichotomously scored individual questions. When the cluster items were broken down by their individual questions, there was a grand total of 179 questions available for analysis (this breakdown will be utilized when compiling the data for

Research Questions 3 and 4). The distribution of items by content strand and ability level is shown in Table 3.

Procedure for Assessment Administration

A brief informational description of the assessment administration is provided in this section in the following paragraphs. For complete details of the assessment procedure see Chapter 5 of The *NAEP 1996 Technical Report* (National Center for Education Statistics, 1999). Test administration was conducted at the selected schools using local exercise administrators who were responsible for carrying out the assessments in accordance within established NAEP protocols. All assessments were administered between January 3 and March 29, 1996. Each session proceeded as follows:

1. Students selected for the assessment reported to the designated testing room.
2. Exercise administrators read aloud a script describing the assessment.
3. Assessment booklets were distributed.
4. Additional scripted directions were read (for students who received a block of calculator items this is the point when they were provided with a calculator).
5. Students began taking the assessment.

Exercise administrators monitored the room during assessments to insure that students were working on the correct section of their booklet and to discourage cheating.

Table 3

Distribution of Items by Content Strand and Ability Level

Content strand	Ability level			Grand total
	Conceptual understanding	Procedural knowledge	Problem solving	
Number sense, properties, and operations	17 (9%)	16 (9%)	17 (9%)	50 (28%)
Measurement	7 (4%)	9 (5%)	11 (6%)	27 (15%)
Geometry and spatial sense	11 (6%)	3 (2%)	18 (10%)	32 (18%)
Data analysis, statistics and probability	11 (6%)	7 (4%)	15 (8%)	33 (18%)
Algebra and functions	19 (11%)	10 (6%)	8 (4%)	37 (21%)
Grand Total	65 (36%)	45 (25%)	69 (39%)	179 (100%)

Students who received a booklet containing a block of calculator items were provided with a nonprogrammable scientific calculator (i.e., TI 30 Challenger). As students proceeded through the calculator block of questions, each item had a place for students to indicate whether or not they had used the calculator on that particular item.

Analysis

Each research questions is restated below, followed by a detailed explanation of the analysis as it relates to each question. Before proceeding it is important to re-emphasize the unique factors and considerations that must be kept in mind when working with NAEP data. Readers are encouraged to review these factors in the *Delimitations* section of Chapter I.

Question 1: How does frequency of calculator use in the classroom relate to mathematics achievement on the NAEP mathematics assessment?

This question was addressed using the *NAEP Data Tool v. 3.0*. Descriptive statistics (average scale score, standard deviation, and standard error of measure) were computed for all assessed students as a reference point, followed by group results based on student reported frequency of calculator use.

The amount of calculator use was determined using student background questionnaire item M812711 *How often do you use a calculator for math?* Though there are three other background questions concerning how often students use a calculator (M812001-3), these questions confine calculator use to certain specific conditions (e.g., in-class work, homework, and tests/quizzes). Because M812711 was the most general and inclusive background question on calculator use it was chosen as the determining factor to classify subjects into calculator use groups.

Because the NAEP student background questionnaires are filled out by students and subject to self-reporting biases, the student-reported results were cross checked using information from the teacher's background questionnaire item T044505 *How often do the students in this class...use a calculator?* This cross check provided an indication of the reliability of the student-reported information from using a second source. Keep in mind that though the teacher background questionnaire was used to extract the data, the unit of analysis was still the student (i.e., results from the teacher's reported use of calculators were computed using data only from students who could be uniquely matched to their teacher).

Statistically significant differences in score were determined using NAEP standardized procedures (National Center for Education Statistics, 2001, pp. 247-254). These conventions are built in to the *NAEP Data Tool* and run pairwise comparisons using the Benjamini-Hochberg (1994) False Discovery Rate (FDR) to control familywise error (i.e., the inflated type I error associated with multiple comparisons).

Because the large sample size had the potential to be "too powerful" at detecting statistically significant differences, a standardized mean difference effect size was computed to assess the "practical significance" of group differences (Huck, 2000, p. 208).

Question 2: How is this relationship with achievement affected when potentially confounding variables are controlled (e.g., teacher's education and experience, student's socioeconomic status, parent's education level, etc.)?

Other factors that may affect student's scores are addressed in Research Question 2. Specifically, these analyses controlled for the following student factors; *gender, SES* (as indicated by *National School Lunch Program* eligibility [student level data] and the school's *Title I* status [school level data]), p*arents highest level of education*, *student's NAEP mathematics achievement level,* and *type of school* (i.e., *public or nonpublic*). Two teacher factors were also controlled: *teacher's knowledge of the NCTM Curriculum and Evaluation Standards for School Mathematics,* and whether or not the teacher had *studied the use of calculators in mathematics instruction.*

The results of these analyses were determined for each calculator use group. General trends in score were compared with the results from *Research Question 1* as well as how those trends hold within the controlled factor.

Question 3: How is this relationship affected when the data are disaggregated by question-type, where the calculator is allowed on some NAEP questions but not others?

This question required item level analysis and presented two analytical problems: (a) *NAEP Data Tool* does not analyze data at the item level, and (b) *NAEP Scale Scores* are not an appropriate outcome measure for item level analysis. These two issues were resolved using raw data from the *NAEP Restricted Use Data* set, but this presented a problem with polytomously scored items and how to deal with partial credit. Dichotomously scored items are straight-forward: the response is either correct or incorrect. Polytomously scored items, on the other hand, could be scored as either full credit, partial credit, or incorrect. The options for resolving this situation were to error on the lenient side by counting partial credit as full, or to error on the conservative side by counting only full credit responses. Either way the results would underestimate or overestimate achievement. The decision was made to hold to the higher standard and count only responses that received full credit.

This analysis was performed by converting the raw data results to percent correct and ordinal rank finish (i.e., 1^{st}-4^{th}) formats for each item by calculator use category. An example of the data in these formats is illustrated in Table 4.

Table 4

Example of Results for Selected Questions in Percent Correct and Rank Finish Format

	Frequency of calculator use			
Item #	Almost every day	Once or twice a week	Once or twice a month	Never or hardly ever
Percent correct format				
M070001L	80.96	80.97	79.87	65.29
M072801N	63.55	55.77	54.79	52.94
M072901N	80.65	74.69	75.12	66.35
M073001N	67.11	61.34	57.99	52.49
M073101N	55.21	59.04	46.58	51.14
Rank finish format				
M070001L	2	1	3	4
M072801N	1	2	3	4
M072901N	1	3	2	4
M073001N	1	2	3	4
M073101N	2	1	4	3

The data was analyzed using a two-way ANOVA with *frequency of calculator use* and *calculator allowed/restricted* as the factors. Significance tests were conducted on each factor's main effect and on the interaction effect. To further analyze trends in performance, the data was analyzed by *difficulty level* (NAEP defines difficulties by percentage of correct responses as follows; *easy*: greater than or equal to 60%, *Moderate*: greater than or equal to 40% and less than 60%, *Hard*: less than 40%), followed by content strand and ability level.

The rank-finish data was computed and analyzed using a Friedman's two-way analysis of variance of ranks to determine if rank finishes were equally distributed between calculator use groups.

Question 4: How do groups compare between frequency of calculator use and their ability to accurately recognize when it is appropriate and inappropriate to apply a calculator to solve a problem?

This question was addressed using raw data from the *NAEP Restricted Use Data* files. The criteria of correct application used the same standard established by Mullis et al. (1991), which defined calculator proficiency as follows (p. 178):

> High Group—Students who both 1) indicated that they had used the calculator for at least half of the calculator active items they were presented *and* 2) used the calculator appropriately at least 85%of the time (i.e., used it for the calculator-active items and did not use it for the calculator-inactive items).
>
> Other Group—Students who either 1) indicated that they had used the calculator for less than half of the calculator-active items they were presented *or* 2) did not use the calculator appropriately at least 85%of the time.

The outcome measure for this question is the percentage of students in the *high group* and *other group*.

This concludes the explanation of the methods used for this study. The next chapter presents the results obtained from conducting these analyses.

CHAPTER IV

RESULTS

As stated in Chapter I, the purpose of this study was to analyze the 1996 NAEP data to identify the relationship between calculator use and performance on the NAEP Mathematics assessment. This chapter is organized based on the four research questions posed in Chapter I. For each research question a series of results are provided which includes (a) the descriptive statistics, (b) results of statistical significance tests, and (c) graphs of trends (where appropriate).

Tests of statistical significance show three pieces of information for each comparison:

1. The direction of differences between groups, represented by <, >, or =.
2. The magnitude of differences in mean scale score, represented by *diff* =.

(Note: there may be discrepancies or illogical *diff* values [such as "-0"] between the descriptive table and the statistical significance table. These results are due to rounding error and do not adversely effect the analysis.)

3. The statistical probability (p =) of the observed differences.

To see how one value compares with the others, read across the row for that value.

All statistical tests were conducted using an initial alpha level of $\alpha = 0.05$. The FDR, described by Benjamini and Hochberg (1994), was used to control family-wise error.

Question 1: How does frequency of calculator use in the classroom relate to mathematics achievement on the NAEP mathematics assessment?

Table 5 displays the number of students, average scale score, standard deviation, and standard errors (listed in parentheses) for all assessed students as a baseline for comparison. Following this result the table displays the same information based on student-reported and teacher-reported frequency of calculator use. Results of this analysis from both student-reported and teacher-reported data show a trend indicating that more frequent use of calculators is associated with higher scores.

Before proceeding with any statistical tests of comparison it was important to determine if the statistical assumptions underlying the comparisons had been met, the primary concern being the assumption of homogeneity of variance due to unequal sample sizes between calculator use groups. Table 6 shows the results of the statistical significance test for homogeneity of variance between calculator use groups. Results are not statistically significant and indicate that the assumption of homogeneity of variance between groups cannot be rejected. With this assumption met, the robustness of further statistical tests was no longer a cause for concern.

Results of the statistical significance tests for the student-reported and teacher-reported use of calculators is given in Tables 7 and 8, respectively. In both cases the results indicate that overall, more frequent use of calculators corresponds to higher scores. Though the p values vary between student-reported and teacher-reported data, the trend in scores is consistent for both groups. The difference in between-group scores is statistically significant for all comparisons with the exceptions of the *weekly* versus *monthly* group ($p = 0.44$) from the student-reported data and both the *weekly* versus *monthly* and *never* versus *monthly* ($p = 0.06$ and $p = 0.28$, respectively) groups from the teacher-reported data.

Table 5

Scale Score and Standard Deviation (with Standard Error of Measure in Parentheses) on the NAEP Mathematics Exam for All Students and By Calculator Use Category as Reported by Students (M812711) and Teachers (T044505)

	N	Row percentage	Average scale score	Standard deviation
All students	7033[a,b]	100	272 (1.1)	36 (0.6)
Score by calculator use category (student reported):				
Almost every day	3187[c]	48 (2.3)	280 (1.5)	35 (0.8)
1-2 times a week	1801[c]	26 (1.3)	268 (1.3)	36 (1.1)
1-2 times a month	1084[c]	14 (0.9)	267 (1.8)	34 (1.0)
Never or hardly ever	961[c]	12 (1.0)	258 (2.2)	37 (1.2)
Score by calculator use category (teacher reported):	5752[b]			
Almost every day	--	55 (2.7)	281 (1.7)	--
1-2 times a week	--	21 (2.5)	271 (3.0)	--
1-2 times a month	--	14 (2.1)	263 (3.1)	--
Never or hardly ever	--	9 (1.5)	256 (3.9)	--

Note. The NAEP Mathematics scale ranges from 0 to 500.
[a] This number does not coincide with the $N = 7146$ cited in Table 2. The difference of 113 represents the students who were removed from the analysis because they either omitted or gave multiple responses to the background question.
[b] The discrepancy in *N* size between teachers and students was due to the fact that not all students could be matched to their teacher. The number shown represents the number of students who could be uniquely matched to their teacher.
[c] These values were extracted from the raw data and were not available from the *NAEP Data Tool*.
-- *NAEP Data Tool* did not compute this variable for teacher reported data.

Table 6

Results from the Statistical Significance Test of Homogeneity of Variance

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		= diff = -1 p = 0.5918	= diff = 1 p = 0.6090	= diff = -2 p = 0.0961
1-2 times a week			= diff = 1 p = 0.3600	= diff = -2 p = 0.2963
1-2 times a month				= diff = -3 p = 0.0564

Table 7

Result for Average Scale Score Differences Using Student-Reported Use of Calculators

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		> diff = 11 p = 0.0000	> diff = 13 p = 0.0000	> diff = 22 p = 0.0000
1-2 times a week			= diff = 2 p = 0.4472	> diff = 10 p = 0.0001
1-2 times a month				> diff = 9 p = 0.0027

Table 8

Result for Average Scale Score Differences Using Teacher-Reported Use of Calculators

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		> diff = 10 $p = 0.0058$	> diff = 18 $p = 0.0000$	> diff = 24 $p = 0.0000$
1-2 times a week			= diff = 8 $p = 0.0643$	> diff = 14 $p = 0.0054$
1-2 times a month				= diff = 6 $p = 0.2176$

The standardized mean difference effect size (*d*) of the between-group comparisons is given in Table 9 and indicates larger effects with more frequent calculator use. Note that because the results from the student-reported and teacher-reported data were consistent with each other, all results from this point forward are calculated using student-reported data unless otherwise indicated.

Research Question 1 was addressed a second time using data from the 2000 NAEP. A comparison of the 2000 and 1996 data, provided in Table 10, showed an increase in scores in 2000 for all groups, but the only statistically significant difference between administration years was within the *weekly* and *monthly* groups compared between years (see Table 11). The line graph of this data, shown in Figure 1, did not indicate an interaction effect.

The 2000 results, listed in Table 12, showed the same pattern of higher scores associated with more frequent calculator use found in the 1996 data. The difference in scores between calculator use groups from the 2000 data was statistically significant for all comparisons except the *weekly* versus *monthly* groups, just as it was in 1996.

Table 9

Standardized Mean Difference Effect Size by Frequency of Calculator Use (Never or Hardly Ever Is Comparison Group)

Effect size	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
d	0.61	0.29	0.24	---

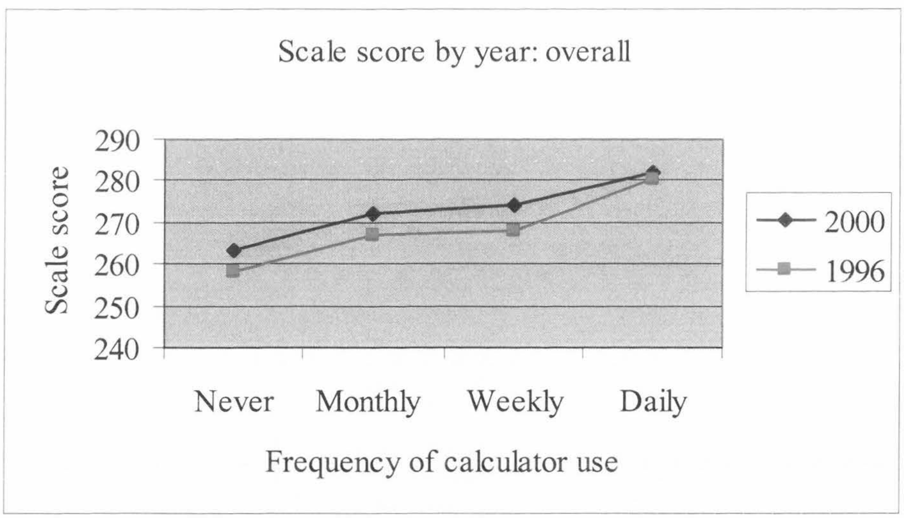

Figure 1. Overall scores: 2000 versus 1996.

Table 10

Average Scale Score and Row Percentage (with Standard Error in Parentheses) for Years 2000 and 1996

		Almost every day		1-2 times a week		1-2 times a month		Never or hardly ever	
Year	*N*	Average scale score	Row percentage	Average scale score	Row percentage	Average scale score	Row percentage	Average scale score	Row percentage
2000	15,464	282 (1.1)	48 (1.4)	274 (0.9)	25 (0.7)	272 (1.3)	13 (0.7)	263 (1.5)	13 (0.9)
1996	7,033	280 (1.5)	48 (2.3)	268 (1.3)	26 (1.3)	267 (1.8)	14 (0.9)	258 (2.2)	12 (1.0)

Table 11

Statistical Significance of Differences in Overall Score Between Years 2000 and 1996 by Calculator Use Category

Years	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
2000 versus 1996	= diff = 2 $p = 0.2494$	> diff = 5 $p = 0.0013$	> diff = 5 $p = 0.0179$	= diff = 5 $p = 0.0504$

Table 12

Result for Average Scale Score Differences in 2000 Using Student-Reported Use of Calculators

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		> diff = 8 p = 0.0000	> diff = 10 p = 0.0000	> diff = 19 p = 0.0000
1-2 times a week			= diff = 2 p = 0.3223	> diff = 10 p = 0.0000
1-2 times a month				> diff = 9 p = 0.0000

Question 2: How is this relationship with achievement affected when potentially confounding variables are controlled (e.g., teacher's education and experience, student's socioeconomic status, parent's education level, etc.)?

Other factors that might affect student's scores are addressed in *Research Question 2*. Specifically, these analyses controlled for the following student factors; *gender, SES* (as indicated by *National School Lunch Program* eligibility [student level data] and the school 's *Title I* status [school level data]), *parents' highest level of education*, *student's NAEP Mathematics achievement level,* and *type of school* (i.e., *public or nonpublic*). Two teacher factors were also controlled: (a) teacher's knowledge of the NCTM Curriculum and Evaluation Standards for School Mathematics, and (b) whether or not the teacher had studied the use of calculators in mathematics instruction. Comparisons between 1996 and 2000 data were consistent with each other and statistically insignificant in all cases except for the teacher variable of studied the use of calculators in mathematics instruction, which will be discussed at the end of this section.

Scale score results based on *gender* are listed in Table 13 with the statistical significance of these results presented in subsequent tables. Results show no significant differences between genders (Table 14). Separate analyses by gender reveal that higher scores are associated with more frequent calculator use and that these differences are statistically significant except for the *weekly* versus *monthly* comparison for males (Table 15) and the *weekly* versus *monthly* and *monthly* versus *never* comparisons for females (Table 16).

Table 13

Average Scale Score and Row Percentage (with Standard Errors in Parentheses) by Student's Gender and Frequency of Calculator Use

	Male $N = 3541$		Female $N = 3492$	
Group	Score	Row percentage	Score	Row percentage
Overall	272 (1.4)	50 ([a])	272 (1.1)	50 ([a])
Calculator use category				
Almost Every Day	280 (1.9)	47 (2.3)	280 (1.5)	50 (2.4)
1-2 a week	268 (1.9)	27 (1.4)	269 (1.6)	25 (1.5)
1-2 a month	268 (2.3)	14 (1.0)	265 (1.8)	13 (1.0)
Never or Hardly Ever	256 (3.0)	12 (1.1)	260 (2.3)	12 (1.1)

[a] Standard error estimates cannot be accurately determined.

Table 14

Statistical Significance of Differences in Gender by Calculator Use Category

		gender	
Frequency of calculator use	gender	male	female
All students	male		= diff = -1 $p = 0.7127$
	female	= diff = 1 $p = 0.7127$	
Almost every day	male		= diff = 0 $p = 0.9567$
	female	= diff = 0 $p = 0.9567$	
1-2 times a week	male		= diff = 0 $p = 0.8597$
	female	= diff = 0 $p = 0.8597$	
1-2 times a month	male		= diff = 3 $p = 0.2656$
	female	= diff = -3 $p = 0.2656$	
Never or Hardly Ever	male		= diff = -3 $p = 0.3622$
	female	= diff = 3 $p = 0.3622$	

Table 15

Statistical Significance of Differences by Calculator Use Category Within Gender = Male

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		> diff = 12 $p = 0.0001$	> diff = 11 $p = 0.0003$	> diff = 23 $p = 0.0000$
1-2 times a week			= diff = -0 $p = 0.9902$	> diff = 12 $p = 0.0014$
1-2 times a month				> diff = 12 $p = 0.0024$

Table 16

Statistical Significance of Differences by Calculator Use Category Within Gender = Female

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		> diff = 11 $p = 0.0000$	> diff = 15 $p = 0.0000$	> diff = 20 $p = 0.0000$
1-2 times a week			= diff = 4 $p = 0.1349$	> diff = 9 $p = 0.0023$
1-2 times a month				= diff = 5 $p = 0.0887$

Socioeconomic status can be inferred from the NAEP data using available information at both the individual and the school levels. Analyzing the data based on National School Lunch Program Eligibility serves as an indicator of SES at the individual level; results of this analysis are presented in Table 17. Subsequent tables show the statistical significance of these results for National School Lunch Program eligible (Table 18) and noneligible students (Table 19). Results show significant differences favoring calculator use for both groups, but the differences are smaller for eligible students than noneligible students. A second way of estimating SES is to use the school's Title I status. Using Title I status takes into account SES on a schoolwide basis, rather than an individual basis. Table 20 shows the scale score results based on the school's Title I status. The results indicate that schools designated as Title I participants have lower scores than non-Title I schools. These results continue to show the trend of higher scores associated with more frequent calculator use within Title I schools, but the differences are far less dramatic than those found in prior analyses. Furthermore, there is a three point negative effect associated with calculator use between the *weekly* and *monthly* categories within Title I schools. As shown in Table 21, none of the calculator use group comparisons within Title I eligible schools are statistically significant.

Schools that do not participate in Title I display a trend of higher scores associated with more frequent calculator use. The results of group comparisons for these schools, shown in Table 22, were all statistically significant with the exception of the *weekly* versus *monthly* comparison.

These results indicate that after controlling for SES, higher scores are associated with more frequent calculator use. This trend is consistent with earlier results, but the statistical significance is less dramatic, particularly for the Title I eligible students.

Parents' level of education was the next variable to be controlled. Table 23 displays the scale score results by calculator use category and parents' level of education. Subsequent tables show the statistical significance of the data based on parents' level of education using the following categories: less than high school (Table 24), graduated high school (Table 25), some education after high school (Table 26), graduated college (Table 26), and unknown (Table 28).

As might be expected, higher levels of parents' education were associated with higher scores on the NAEP assessment. However, the results also indicate that higher scores were associated with more frequent calculator use regardless of parents' level of education with only one exception—*weekly* versus *monthly* within the *unknown* level of education.

The statistical significance of differences varies within this control factor. For less than high school none of the differences are statistically significant (Table 24). For graduated high school (Table 25) and unknown (Table 28) the only statistically significant difference is between the *daily* and *never* groups, with results favoring the *daily* group. Table 28 also reveals a slightly negative calculator effect for the *weekly* and *monthly* comparison, but the difference is not statistically significant. For the some education after high school category, four of six comparisons favored the more

Table 17

Average Scale Score and Row Percentage (with Standard Errors in Parentheses) by National School Lunch Program Eligibility

National School Lunch Program eligibility	*N*	Almost every day				1-2 times a week				1-2 times a month				Never or hardly ever			
		Average scale score		Row percentage		Average scale score		Row percentage		Average scale score		Row percentage		Average scale score		Row percentage	
All examinees	7033	280	(1.5)	48	(2.3)	268	(1.3)	26	(1.3)	267	(1.8)	14	(0.9)	258	(2.2)	12	(1.0)
Eligible	1805	257	(2.2)	40	(2.3)	249	(2.3)	28	(1.3)	252	(2.2)	17	(1.1)	243	(2.7)	15	(1.5)
Non eligible	3876	286	(1.8)	49	(3.2)	277	(1.6)	27	(1.9)	276	(2.2)	13	(1.2)	265	(3.0)	11	(1.3)

[a] Data not available for all students

Table 18

Statistical Significance of Differences by Calculator Use Category Within National School Lunch Program Eligible Students

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		> diff = 9 p = 0.0067	= diff = 5 p = 0.0994	> diff = 14 p = 0.0002
1-2 times a week			= diff = -4 p = 0.2555	= diff = 5 p = 0.1528
1-2 times a month				> diff = 9 p = 0.0149

Table 19

Statistical Significance of Differences by Calculator Use Category Within National School Lunch Program Noneligible Students

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		> diff = 8 p = 0.0017	> diff = 10 p = 0.0013	> diff = 20 p = 0.0000
1-2 times a week			= diff = 2 p = 0.5685	> diff = 12 p = 0.0010
1-2 times a month				> diff = 10 p = 0.0075

frequent use of calculators (Table 26). For the *graduated college* level, the results for the *daily* users are statistically significant, but comparisons among the other three remaining groups are not statistically significant (Table 27).

The next variable to be controlled was the student's NAEP achievement level. This analysis was run to determine the effect of calculators within achievement classifications. The reader is reminded of the NAEP achievement levels and their score ranges: below basic: 0 to 261; basic: 262 to 298; proficient: 299 to 332; advanced: 333 to 500 (National Center for Education Statistics, 1999, p. 34).

One caution to keep in mind when looking at achievement level data is that by definition it is disaggregated such that there will be differences between achievement levels (i.e., below basic, basic, proficient, and advanced), but not necessarily within those achievement levels (i.e., differences between *daily*, *weekly*, *monthly* and *never* users within the same achievement level). After controlling for achievement level, the results show a trend of higher scores associated with more frequent calculator use. These results are shown in Table 29 and Figure 2.

Statistical significance tests for the advanced achievement level were not possible due to the fact that *weekly*, *monthly*, and *never* calculator use groups within this level were too small to permit reliable estimates. This fact is worth noting as the vast majority (68%) of students scoring at this level report using calculators on a *daily* basis. Combining the *daily* and *weekly* users accounts for 88% of scores at this level, while a mere 5% of students from the *never* category score at the advanced level.

Table 20

Average Scale Score and Row Percentage (with Standard Errors in Parentheses) by School's Title I Status

		Almost every day				1-2 times a week				1-2 times a month				Never or hardly ever			
School's Title I status	*N*	Average scale score		Row percentage		Average scale score		Row percentage		Average scale score		Row percentage		Average scale score		Row percentage	
All examinees	7033	280	(1.5)	48	(2.3)	268	(1.3)	26	(1.3)	267	(1.8)	14	(0.9)	258	(2.2)	12	(1.0)
Participated	769	246	(4.8)	36	(3.2)	243	(3.7)	31	(2.5)	246	(4.2)	19	(1.7)	241	(4.8)[a]	14	(2.3)[a]
Did Not Participate	6264	283	(1.4)	50	(2.5)	272	(1.5)	26	(1.4)	271	(1.8)	13	(0.9)	261	(2.5)	11	(1.1)

[a] The nature of the sample does not allow accurate determination of the variability of the statistic.

Table 21

Statistical Significance of Differences for Title I Participating Schools

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		= diff = 4 $p = 0.5678$	= diff = 1 $p = 0.9360$	= diff = 6 $p = 0.4247$
1-2 times a week			= diff = -3 $p = 0.5949$	= diff = 2 $p = 0.7409$
1-2 times a month				= diff = 5 $p = 0.4392$

Table 22

Statistical Significance of Differences for Schools That Did Not Participate in Title I Programs

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		> diff = 10 $p = 0.0000$	> diff = 12 $p = 0.0000$	> diff = 22 $p = 0.0000$
1-2 times a week			= diff = 2 $p = 0.4443$	> diff = 12 $p = 0.0003$
1-2 times a month				> diff = 10 $p = 0.0029$

Results at the proficient and basic achievement levels, shown in Tables 30 and 31 respectively, were not statistically significant between calculator use categories. The below basic results show an increase in scores with more frequent calculator use (Table 32), but the only statistically significant result was between the *daily* versus *never* comparison and favored the *daily* users.

The lines on the graph in Figure 2 show trends by achievement level that are not fully exposed by the overall results found in Research Question 1. Specifically, at the advanced level the trend line reduces to a single data point due to the fact that overwhelming majority students who scored at this level use a calculator on a *daily* basis. The lines representing the proficient and basic achievement levels are essentially flat and indicate little difference in score based on calculator use. The below basic line shows a trend of higher scores associated with more frequent calculator use similar to the slope of the overall results line, but its slope is not as steep and rises only 8 points compared to 22 in the overall line.

The next analysis controlled for the type of school attended; that is, public or nonpublic. The results, shown in Table 33, indicated that regardless of school type, more frequent calculator use corresponded with higher scores.

Significance tests within public schools are shown in Table 34 and indicated statistically significant differences favoring calculator use for all comparisons except *weekly* versus *monthly*. Results for nonpublic schools are shown in Table 35 and

Table 23

Average Scale Score and Row Percentage (with Standard Errors in Parentheses) by Parents' Level of Education

Parents' highest level of education	*N*	Almost every day				1-2 times a week				1-2 times a month				Never or hardly ever			
		Average scale score		Row percentage		Average scale score		Row percentage		Average scale score		Row percentage		Average scale score		Row percentage	
All examinees	7033	280	(1.5)	48	(2.3)	268	(1.3)	26	(1.3)	267	(1.8)	14	(0.9)	258	(2.2)	12	(1.0)
Less Than H.S.	456	260	(2.9)	38	(3.5)	253	(4.0)	27	(2.6)	252	(4.8)	16	(1.6)	249	(4.4)	19	(3.0)
Graduated H.S.	1483	266	(1.6)	43	(2.5)	262	(2.1)	28	(1.8)	259	(3.7)	15	(1.2)	254	(2.9)	14	(1.5)
Some after H.S.	1293	285	(1.9)	49	(2.9)	277	(2.2)	27	(1.9)	272	(3.1)	14	(1.6)	267	(2.5)	9	(1.0)
Graduated college	3074	289	(2.0)	55	(2.9)	276	(2.0)	24	(1.6)	276	(2.3)	12	(1.2)	268	(3.3)	9	(0.9)
Unknown	717	261	(3.0)	39	(2.7)	251	(3.0)	29	(2.4)	256	(4.4)	15	(1.6)	244	(4.9)	17	(1.8)

Table 24

Statistical Significance of Differences by Calculator Use Category and Parents' Highest Level of Education Equal to "Less Than High School"

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		= diff =7 $p = 0.1547$	= diff =7 $p = 0.2027$	= diff =10 $p = 0.0557$
1-2 times a week			= diff =0 $p = 0.9596$	= diff =3 $p = 0.5882$
1-2 times a month				= diff =3 $p = 0.6582$

Table 25

Statistical Significance of Differences by Calculator Use Category and Parents' Highest Level of Education Equal to "Graduated High School"

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		= diff = 4 $p = 0.1361$	= diff = 7 $p = 0.0920$	> diff = 12 $p = 0.0013$
1-2 times a week			= diff = 3 $p = 0.4914$	= diff = 8 $p = 0.0400$
1-2 times a month				= diff = 5 $p = 0.3316$

Table 26

Statistical Significance of Differences by Calculator Use Category and Parents' Highest Level of Education Equal to "Some Education After High School"

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		> diff = 8 $p = 0.0102$	> diff = 12 $p = 0.0017$	> diff = 18 $p = 0.0000$
1-2 times a week			= diff = 5 $p = 0.2189$	> diff = 10 $p = 0.0029$
1-2 times a month				= diff = 6 $p = 0.1716$

Table 27

Statistical Significance of Differences by Calculator Use Category and Parents' Highest Level of Education Equal to "Graduated College"

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		> diff = 13 $p = 0.0000$	> diff = 13 $p = 0.0001$	> diff = 22 $p = 0.0000$
1-2 times a week			= diff = 0 $p = 0.8788$	= diff = 9 $p = 0.0261$
1-2 times a month				= diff = 8 $p = 0.0407$

Table 28

Statistical Significance of Differences by Calculator Use Category and Parents' Highest Level of Education Equal to "Unknown"

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		= diff = 10 $p = 0.0241$	= diff = 5 $p = 0.3459$	> diff = 17 $p = 0.0046$
1-2 times a week			= diff = -5 $p = 0.3968$	= diff = 7 $p = 0.2026$
1-2 times a month				= diff = 12 $p = 0.0764$

indicate that statistically significant differences exist between the *daily* users and all subsequent categories, while results within the *weekly*, *monthly* and *never* group comparisons are nonsignificant.

The final control variables accounted for the teacher's knowledge of the *NCTM Curriculum and Evaluation Standards for School Mathematics* and their training in the use of calculators. Table 36 displays the scale score results based on the teacher's reported knowledge level of the *NCTM Standards*. Results suggest that greater knowledge of the *Standards* is associated with higher student achievement. The statistical significance of these results are provided in Table 37 and reveal that reported differences of at least two rank levels are statistically significant, but juxtaposed ranks are nonsignificant.

The other teacher factor considered was whether the teacher had "ever studied (the) use of calculators in mathematics instruction, either in college or university courses or in professional development workshops or seminars." Results of this analysis are provided in Table 38 and the statistical significance of the results in Table 39. All comparisons based on this factor are nonsignificant.

This factor was also run using the 2000 NAEP data and was the one comparison that did show significant results. Between 1996 and 2000 the scores for teachers who responded "yes" to this question had a 4 point increase in score, while those who responded "no" had a 1 point decrease. These differences were statistically significant between years for "yes" responses ($p = 0.0175$), but nonsignificant for "no" responses ($p = 0.8166$) and are illustrated in Figure 3. Additionally, results were statistically significant within the year 2000 ($p = 0.0015$), with those responding "yes" scoring 7 points higher than those responding "no."

Question 3: How is this relationship affected when the data are disaggregated by question-type, where the calculator is allowed on some NAEP questions but not others?

Table 43 shows the percent of correct responses by the item's calculator designation (i.e., allowed or restricted) for all students combined and by frequency of calculator use groups. The group information is graphically presented in Figure 4. These results associate more frequent calculator use with a higher percentage of correct responses for both the calculator allowed and calculator restricted items.

Table 29

Average Scale Score and Row Percentage (with Standard Errors in Parentheses) by Achievement Level

NAEP Achievement Level	*N*	Almost every day		1-2 times a week		1-2 times a month		Never or hardly ever	
		Average scale score	Row percentage	Average scale score	Row percentage	Average scale score	Row percentage	Average scale score	Row percentage
All examinees	7033	280 (1.5)	48 (2.3)	268 (1.3)	26 (1.3)	267 (1.8)	14 (0.9)	258 (2.2)	12 (1.0)
Advanced	267	344 (1.3)	68 (6.5)	---- (---)	20 (5.2)[a]	---- (---)	8 (3.6)[a]	---- (---)	5 (2.1)[a]
Proficient	1439	313 (0.7)	61 (3.4)	313 (0.7)	22 (2.2)	312 (1.3)	10 (1.4)	312 (1.4)	7 (1.2)
Basic	2740	281 (0.4)	50 (2.6)	280 (0.6)	26 (1.4)	280 (0.8)	14 (1.3)	279 (0.9)	10 (1.2)
Below basic	2587	238 (1.0)	38 (2.1)	234 (1.1)	29 (1.5)	235 (1.3)	16 (1.0)	231 (1.5)	17 (1.5)

--- Sample size is insufficient to permit a reliable estimate.

[a] The nature of the sample does not allow accurate determination of the variability of the statistic.

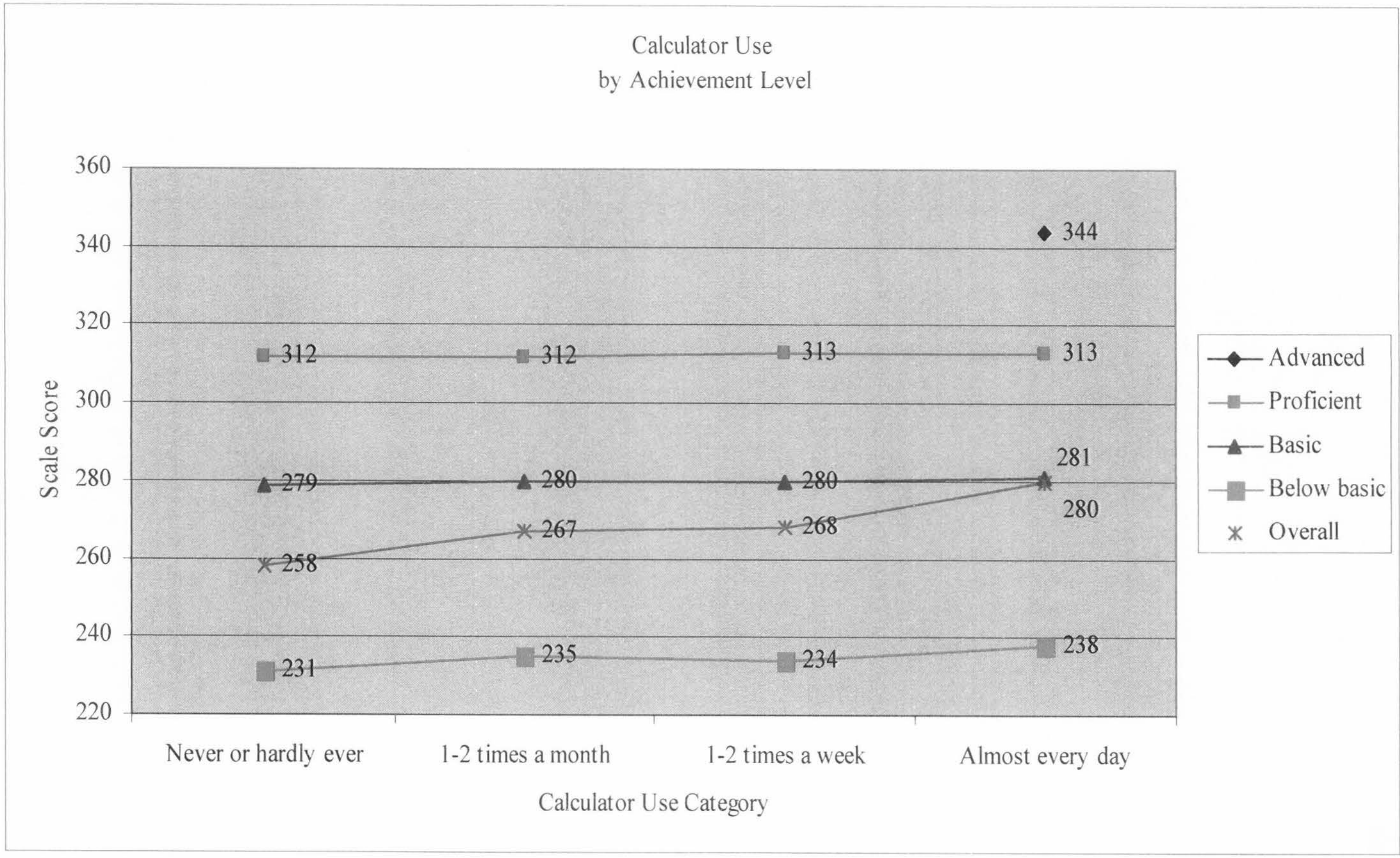

Figure 2. Interaction plot of scale score by calculator use category within NAEP achievement level.

Table 30

Statistical Significance of Differences by Calculator Use Category at the Proficient Level

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		= diff = 0 $p = 0.6891$	= diff = 1 $p = 0.4228$	= diff = 1 $p = 0.3649$
1-2 times a week			= diff = 1 $p = 0.5933$	= diff = 1 $p = 0.5167$
1-2 times a month				= diff = 0 $p = 0.9086$

Table 31

Statistical Significance of Differences by Calculator Use Category at the Basic Level

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		= diff = 1 $p = 0.3502$	= diff = 1 $p = 0.3597$	= diff = 2 $p = 0.1510$
1-2 times a week			= diff = 0 $p = 0.8843$	= diff = 1 $p = 0.4501$
1-2 times a month				= diff = 1 $p = 0.5667$

Table 32

Statistical Significance of Differences by Calculator Use Category at the Below Basic Level

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		= diff = 3 $p = 0.0284$	= diff = 2 $p = 0.1560$	> diff = 7 $p = 0.0003$
1-2 times a week			= diff = -1 $p = 0.5745$	= diff = 3 $p = 0.0699$
1-2 times a month				= diff = 4 $p = 0.0289$

Table 33

Average Scale Score and Row Percentage (with Standard Errors in Parentheses) for Public and Nonpublic Schools

Type of school	*N*	Almost every day				1-2 times a week				1-2 times a month				Never or hardly ever			
		Average scale score		Row percentage		Average scale score		Row percentage		Average scale score		Row percentage		Average scale score		Row percentage	
All examinees	7033	280	(1.5)	48	(2.3)	268	(1.3)	26	(1.3)	267	(1.8)	14	(0.9)	258	(2.2)	12	(1.0)
Public	5492	278	(1.6)	49	(2.5)	267	(1.5)	27	(1.4)	265	(2.0)	13	(0.9)	254	(2.7)	11	(1.1)
Nonpublic	1541	291	(3.1)	43	(5.4)	279	(3.4)	22	(1.9)	280	(3.3)	17	(2.6)	277	(2.8)	18	(2.9)

Table 34

Statistical Significance of Differences by Calculator Use Category Within Public Schools

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		> diff = 11 $p = 0.0000$	> diff = 14 $p = 0.0000$	> diff = 24 $p = 0.0000$
1-2 times a week			= diff = 3 $p = 0.2965$	> diff = 13 $p = 0.0001$
1-2 times a month				> diff = 11 $p = 0.0024$

Table 35

Statistical Significance of Differences by Calculator Use Category Within Nonpublic Schools

Frequency of calculator use	Almost every day	1-2 times a week	1-2 times a month	Never or hardly ever
Almost every day		> diff = 12 $p = 0.0137$	> diff = 12 $p = 0.0140$	> diff = 14 $p = 0.0027$
1-2 times a week			= diff = -0 $p = 0.9621$	= diff = 2 $p = 0.6297$
1-2 times a month				= diff = 2 $p = 0.5881$

Table 36

Scores Based On Teachers Reported Knowledge of the NCTM Curriculum and Evaluation Standards for School Mathematics (Standard Error in Parentheses)

Teacher's reported knowledge of *NCTM Curriculum and Evaluation Standards for School Mathematics*	*N*	Row percentage	Average scale score
All examinees	6030	100	272 (1.1)
Very knowledgeable		16 (2.4)	282 (2.2)
Knowledgeable		32 (3.5)	276 (2.1)
Somewhat knowledgeable		33 (2.9)	270 (2.7)
Little/no knowledge		19 (2.4)	267 (2.3)

Table 37

Statistical Significance of Differences Based on Teacher's Reported Knowledge of the NCTM Curriculum and Evaluation Standards for School Mathematics

Frequency of calculator use	Very knowledgeable	Knowledgeable	Somewhat knowledgeable	Little/no knowledge
Very knowledgeable		= diff = 6 $p = 0.0436$	> diff = 12 $p = 0.0009$	> diff = 15 $p = 0.0000$
Knowledgeable			= diff = 6 $p = 0.0772$	> diff = 9 $p = 0.0055$
Somewhat knowledgeable				= diff = 3 $p = 0.4085$

Table 38

Teachers Reported Training in the Use of Calculators in Mathematics Instruction (Standard Error in Parentheses)

Studied use of calculators in mathematics instruction?		Yes				No			
	N	Average scale score		Row percentage		Average scale score		Row percentage	
	6065	274	(1.4)	78	(2.5)	272`	(2.5)	22	(2.5)

Table 39

Statistical Significance of Differences Based on Teachers' Reported Training in the Use of Calculators in Mathematics Instruction from 1996 Data

Studied use of calculators in mathematics instruction?	Yes	No
Yes		= diff = 2 p = 0.4041
No	= diff = -2 p = 0.4041	

Table 40

Teachers' Reported Training in the Use of Calculators in Mathematics Instruction for Years 1996 and 2000

Studied use of calculators in mathematics instruction?		No				Yes			
Year	*N*	Average Scale Score		Row Percentage		Average Scale Score		Row Percentage	
2000	13,153	271	(1.9)	19%	(1.6)	278	(0.9)	81%	(1.6)
1996	6,065	272	(2.5)	22%	(2.5)	274	(1.4)	78%	(2.5)

Table 41

Statistical Significance of Differences Based on Teachers' Reported Training in the Use of Calculators in Mathematics Instruction from the 2000 Data

Studied use of calculators in mathematics instruction?	Yes	No
Yes		> diff = 7 $p = 0.0015$
No	< diff = -7 $p = 0.0015$	

Table 42

Statistical Significance of Differences Based on Teachers' Reported Training in the Use of Calculators in Mathematics Instruction Between Years 1996 and 2000

Studied use of calculators in mathematics instruction?	Yes	No
2000 versus 1996	> diff = 4 $p = 0.0175$	= diff = -1 $p = 0.8166$

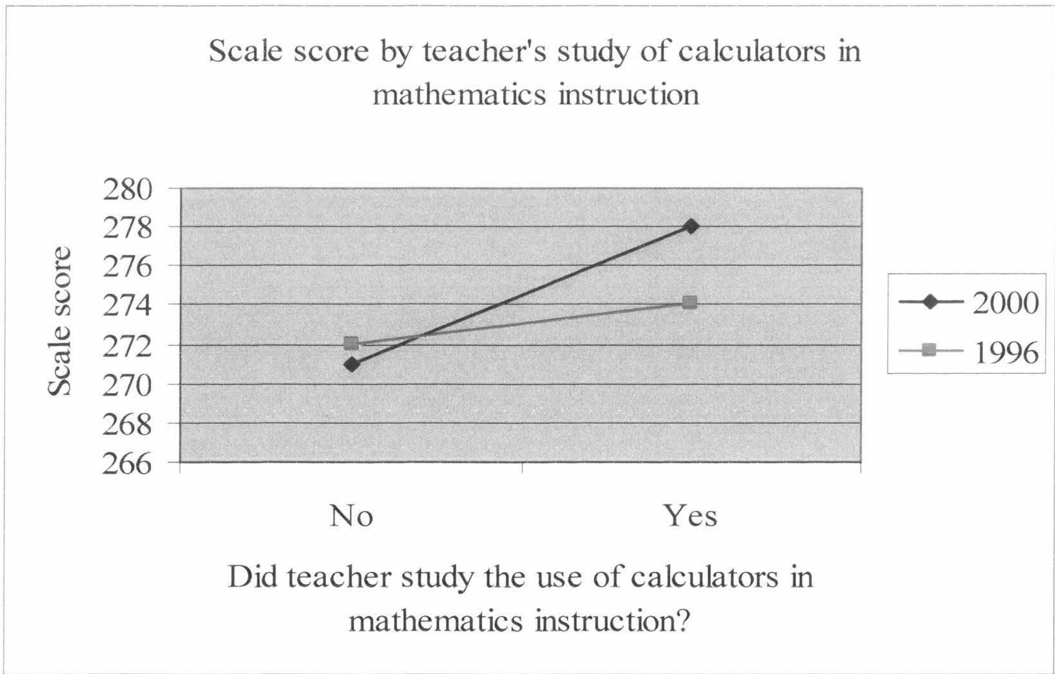

Figure 3. Scale score by teachers' study of calculators in mathematics instruction between years 1996 and 2000.

The statistical significance of this data is displayed in Table 44 and indicates that there are significant differences in main effects based on the item's calculator designation ($p = 0.000$) and the students' frequency of calculator use ($p = 0.000$), but the interaction effect is nonsignificant ($p = 0.195$). Graphical representations of these results are illustrated in Figures 5 and 6.

With this significant finding several other comparisons were conducted to further examine this issue. Table 45 presents the results by item difficulty, and graphs the results by calculator-allowed (Figure 7) and calculator-restricted (Figure 8) items. Performance by content strands is illustrated for calculator-allowed and calculator-restricted items in Figure 9 and Figure 10, respectively. Ability level results are depicted in Figure 11 for calculator-allowed items and Figure 12 for calculator-restricted items. Results from these analyses consistently point to a higher percentage of correct responses with more frequent calculator use regardless of whether the item allowed or restricted the use of a calculator.

Findings from the rank data are given in Table 46 and the statistical significance of results in Table 47. The results show a strong association between more frequent calculator use and higher rank finishes. Results of the Friedman's test are statistically significant for both calculator-allowed and calculator-restricted items.

The effect size results are presented in Table 48. The results indicate that effect sizes become greater as calculator use increases for both calculator-allowed and calculator-restricted items.

Question 4: How does frequency of calculator use in the classroom relate to whether students recognize that it is appropriate or inappropriate to use a calculator to solve specific NAEP problems?

For this question the calculator allowed items are analyzed by their NAEP defined calculator appropriateness categories of active, inactive, and neutral. Table 49 sums the number of times the calculator was used on items within these calculator appropriateness categories and the percent of appropriate application. Results are presented for all students followed by disaggregated data based on student's reported frequency of calculator use. The results indicate that more frequent calculator use is associated with more frequent application on calculator appropriate items as well as less frequent application on calculator inappropriate items (i.e., these students know both when to use the calculator and when not to use it).

Being able to recognize when to apply and when to withhold a calculator is one thing, but it does not necessarily mean that those who are more adept at appropriately applying a calculator will also answer the item correctly. In order to assess competence in both applying the calculator and getting the correct answer a series of criteria were established based on work by Mullis et al. (1991). The criteria proceeded as follows: (a) students must have indicated that they used a calculator for at least half of the calculator allowed items and (b) students had to appropriately apply/withhold calculator use on at least 85% of the calculator allowed items. Those who met these two criteria were qualified as the "high" group, while those who did not were listed as "other."

Table 50 shows the number and percentage of NAEP examinees in the high and other groups, along with the percentage of items answered correctly within each frequency of calculator use. Results indicate that *daily* users meet the high qualification 10% more often than the *weekly* users and 19% more often than the *monthly* and *never* users. In addition, the high group consistently outscores the low group in all frequency of use categories, and within the high and other groups more frequent use of the calculator corresponded to a higher percentage of correct responses, especially in the high group. There was only one exception to this and that was within the other category where the *never* users outperformed the *monthly* users by 2.1% percentage points. Finally, the probability of qualifying in the high group and answering the item correctly is .475 for the *daily* users compared to .283 for the *never* users.

Table 43

Percent of Correct Responses by Item's Calculator Classification and Students' Frequency of Calculator Use

Data for all students	Calculator-allowed items	Calculator-restricted items
Items presented	83,984	197,109
Omitted	4,133	5,586
Not reached	3,860	3,505
Multiple response	66	147
Correct responses	35,078	109,076
Incorrect responses	45,064	84,528
% Correct	43.8	56.3
Data by calculator use category		
Almost every day		
Items presented	36,257	88,015
Correct responses	17,083	52,131
% correct	47.1	59.2
1-2 a week		
Items presented	20,721	49,432
Correct responses	8,790	27,289
% correct	42.4	55.2
1-2 a month		
Items presented	12,332	29,797
Correct responses	4,921	16,222
% correct	39.9	54.4
Never or hardly ever		
Items presented	10,814	26,360
Correct responses	4,284	13,434
% correct	39.6	50.9

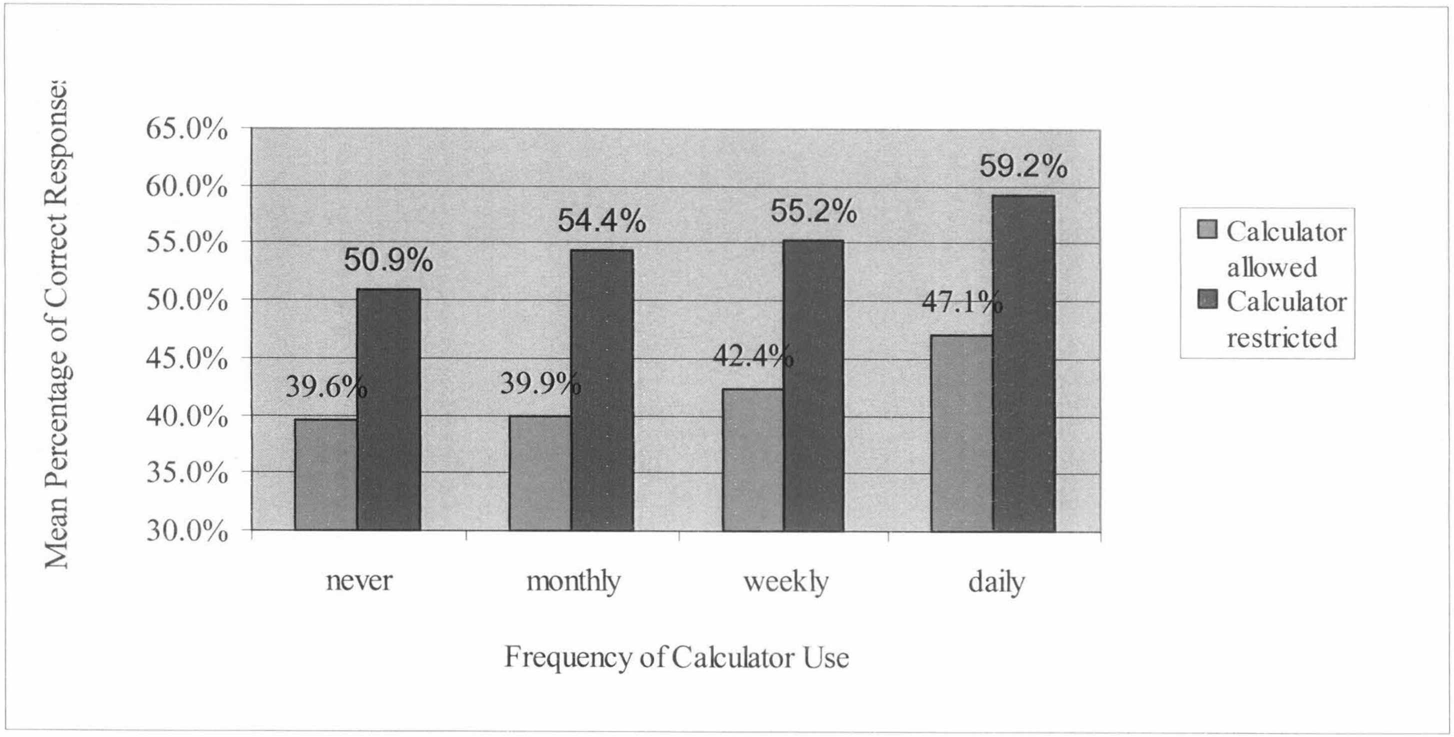

Figure 4. Percent of correct responses on calculator allowed and calculator restricted items by frequency of calculator use.

Table 44

Test of Between Subjects' Effects: Frequency of Calculator Use by Item's Calculator Designation (Calculator-Allowed or Calculator-Restricted)

Source	Type III sum of squares	*df*	Mean square	*F*	Sig.
Corrected model	55.710	7	7.95	199.47	.000
Intercept	2,064.268	1	2,064.26	51,740.06	.000
Frequency of calculator use	10.067	3	3.35	84.10	.000
Item's calculator designation	35.426	1	35.42	887.95	.000
Frequency of calculator use * Item's calculator designation	.188	3	0.06	1.56	.195
Error	472.858	11,852	0.04		
Total	3,478.396	11,860			
Corrected total	528.568	11,859			
Corrected model	55.710	7	7.95	199.47	.000

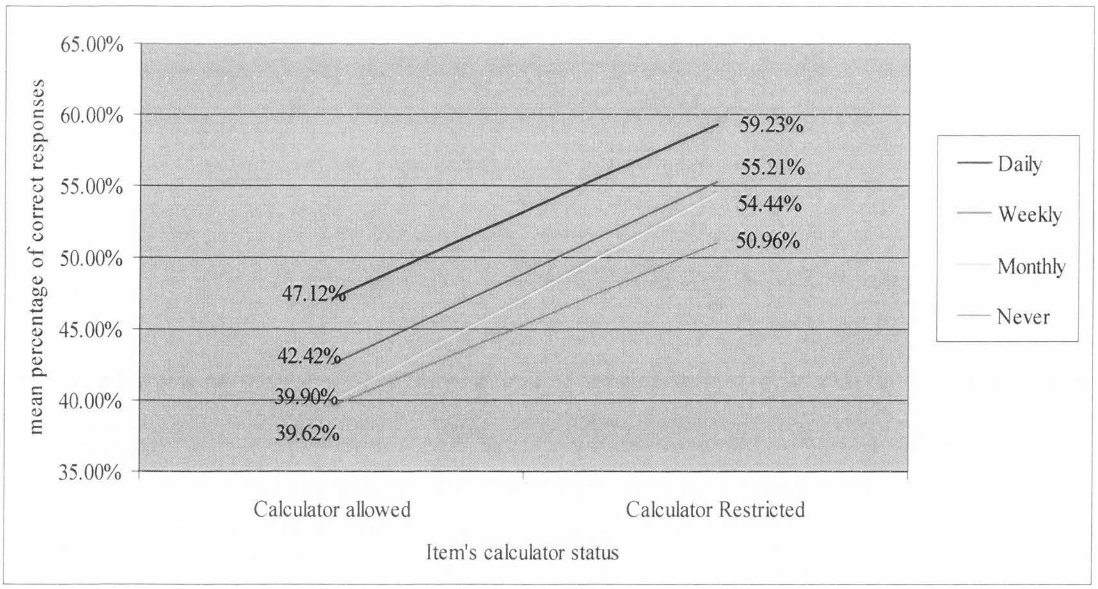

Figure 5. Interaction effect by item's calculator designation.

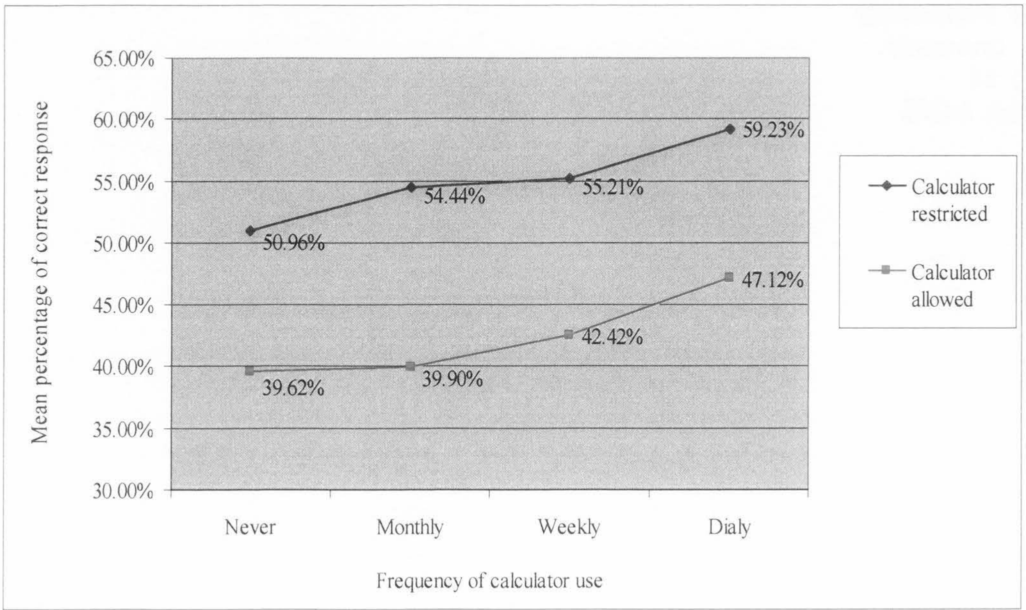

Figure 6. Interaction effects by reported frequency of calculator use.

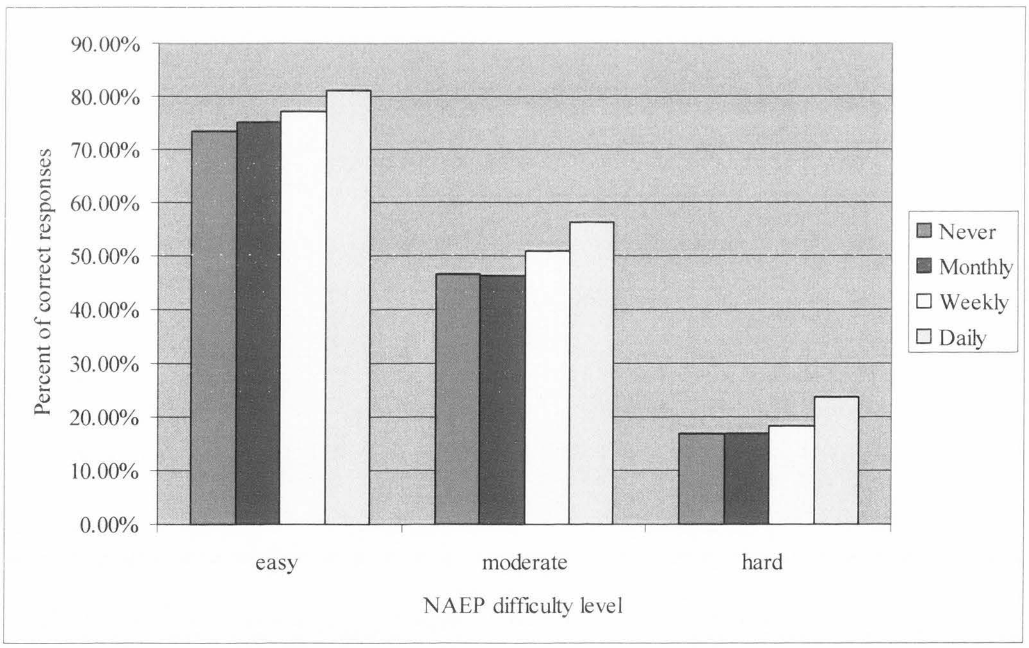

Figure 7. Percent of correct responses by NAEP difficulty level within calculator allowed items.

Table 45

Item Difficulty Within Item's Calculator Designation by Frequency of Calculator Use Category

	Calculator allowed items			Calculator restricted items		
	N = 13	N = 14	N = 26	N = 56	N = 33	N = 33
Difficulty level[a]	easy	moderate	hard	easy	moderate	hard
Frequency of calculator use						
Daily						
Attempted	9,245	9,844	17,168	41,266	23,914	22,835
Correct	7,498	5,531	4,054	32,576	13,261	6,294
% of attempted correct	81.1	56.2	23.6	78.9	55.5	27.6
% above/below *never* group	7.8	9.6	6.7	8.2	11.1	6.0
Weekly						
Attempted	5,335	5,709	9,677	23,204	13,610	12,618
Correct	4,115	2,909	1,766	17,625	6,566	3,098
% of attempted correct	77.1	50.9	18.3	75.9	48.2	24.6
% above/below *never* group	3.8	4.4	1.4	5.2	3.9	3.0
Monthly						
Attempted	3,232	3,282	5,818	14,072	8,100	7,625
Correct	2,426	1,517	978	10,568	3,899	1,755
% of attempted correct	75.1	46.2	16.8	75.1	48.1	23.0
% above/below *never* group	1.8	-0.4	-0.1	4.3	3.8	1.5
Never						
Attempted	2,805	2,956	5,053	12,420	7,184	6,756
Correct	2,056	1,377	851	8,791	3,188	1,455
% of attempted correct	73.3	46.6	16.8	70.8	44.3	21.5
% above/below *never* group	0.0	0.0	0.0	0.0	0.0	0.0

[a] Difficulty levels are defined by NAEP as: Easy: greater than or equal to .60, Moderate: greater than or equal to .40 and less than .60, Hard: less than .40

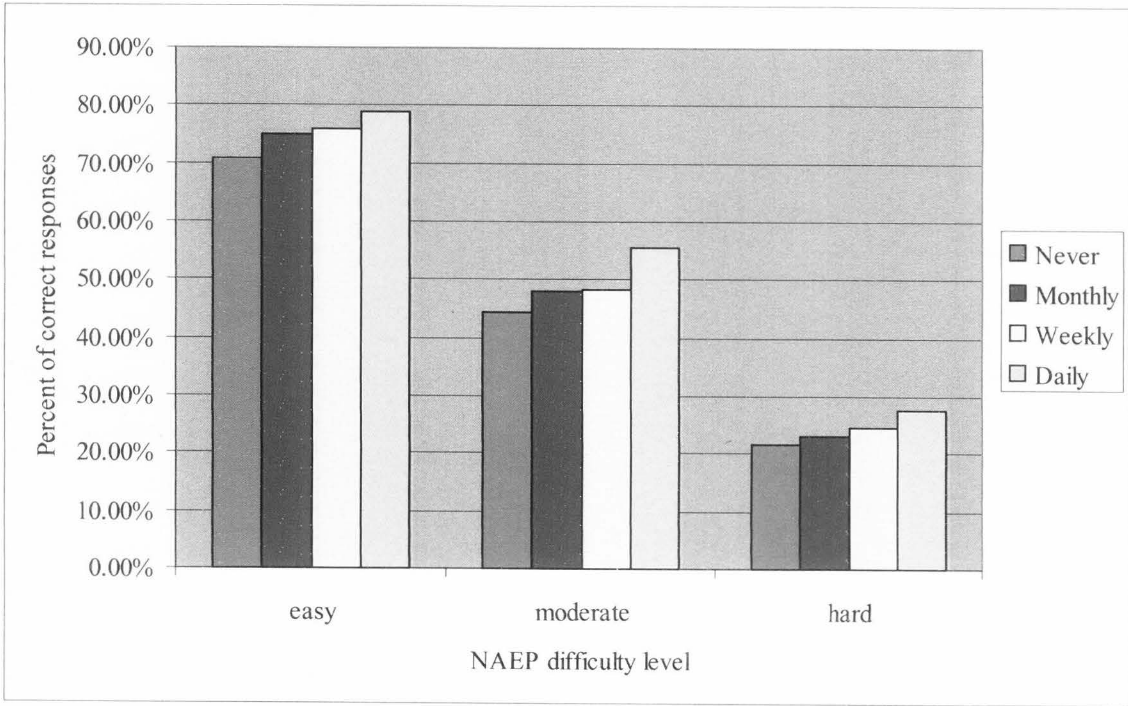

Figure 8. Percent of correct responses by NAEP difficulty level within calculator restricted items.

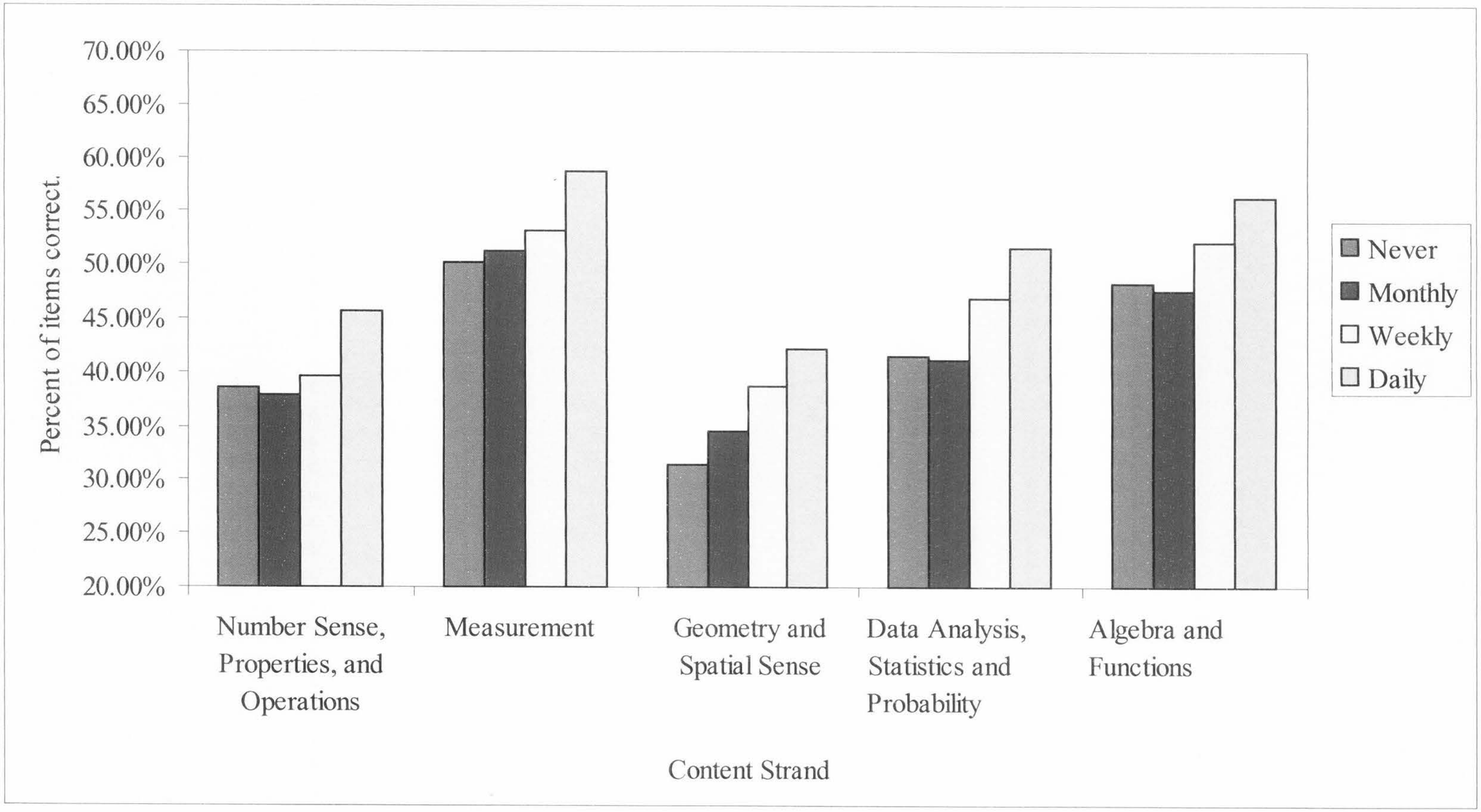

Figure 9. Percent of items correctly answered by content strand and frequency of calculator use for calculator allowed items

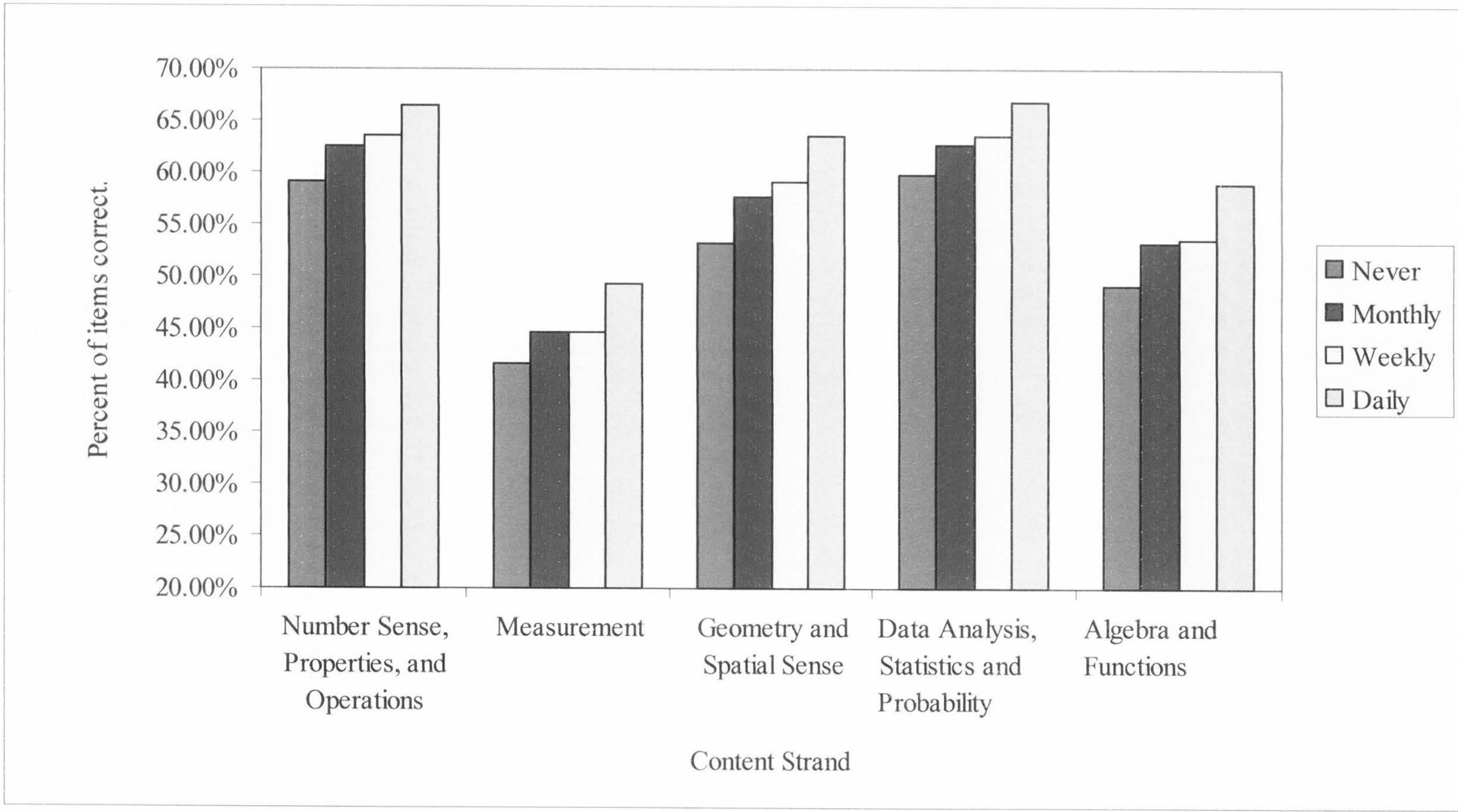

Figure 10. Percent of items correctly answered by content strand and frequency of calculator use for calculator restricted items.

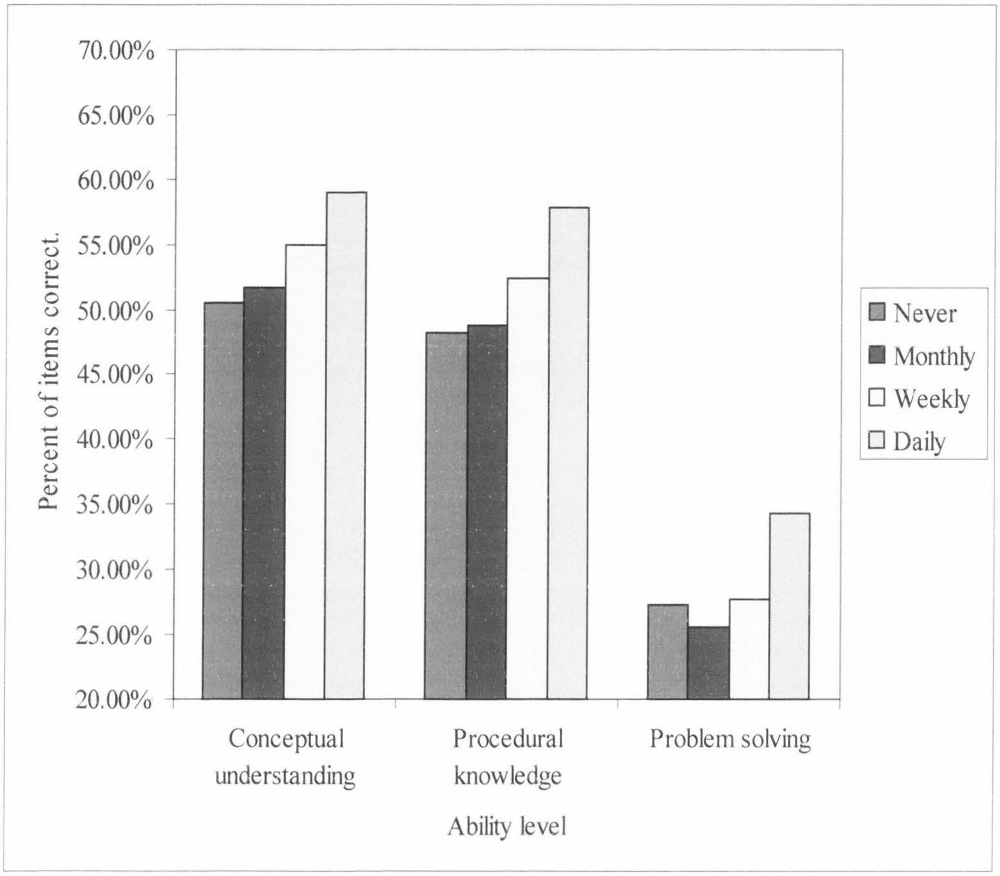

Figure 11. Percent of items correctly answered by ability level and frequency of calculator use for calculator allowed items.

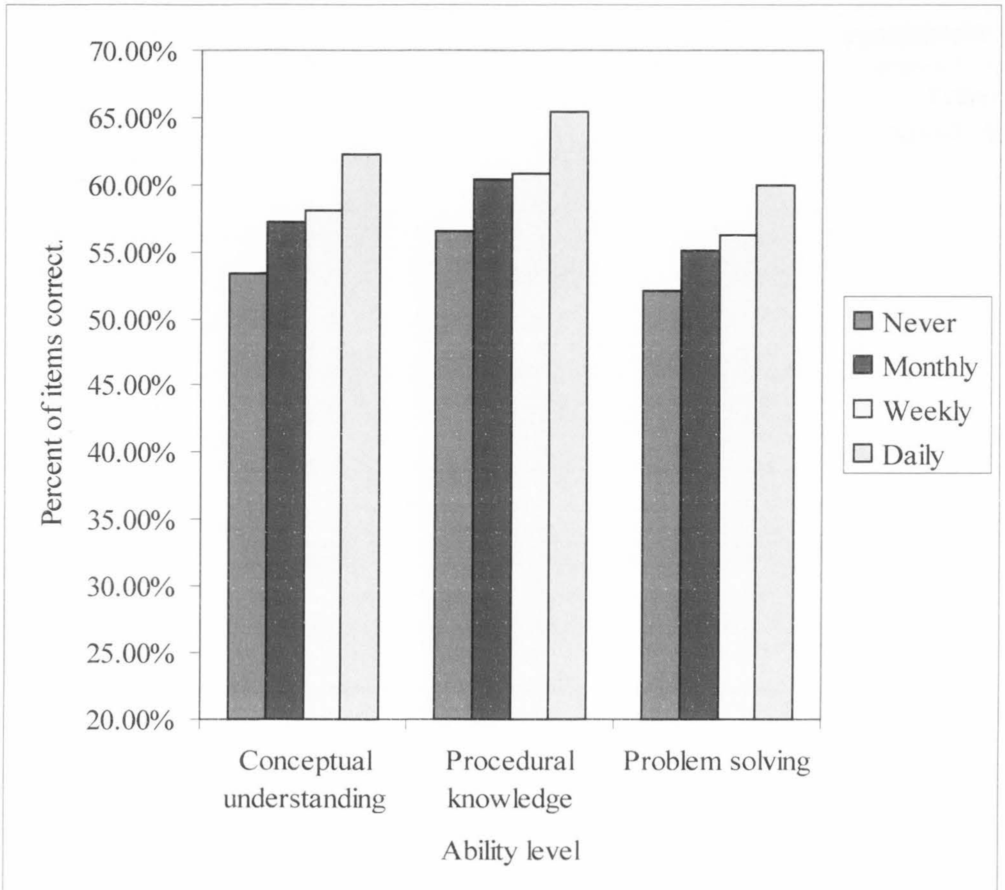

Figure 12. Percent of items correctly answered by ability level and frequency of calculator use for calculator restricted items.

Table 46

Frequency and Percentage (in Parentheses, by Row) of Rank Finish Position by Calculator Use Category

Rank finish position	Almost every day	Once or twice a week	Once or twice a month	Never or hardly ever
1st	142 (82.1 %)	17 (9.8 %)	9 (5.2 %)	5 (2.9 %)
2nd	23 (13.3 %)	88 (50.9 %)	50 (28.9 %)	12 (6.9 %)
3rd	6 (3.5 %)	55 (31.8 %)	78 (45.1 %)	34 (19.7 %)
4th	2 (1.2 %)	13 (7.5 %)	36 (20.8 %)	122 (70.5%)

Table 47

Results of Friedman's Test: Frequency of Calculator Use by Question's Calculator Designation

Item's calculator designation	Mean rank by frequency of calculator use.						
	Daily	Weekly	Monthly	Never	*N*	Chi square	significance
Calculator allowed	1.12	2.38	3.17	3.33	52	95.70	0.000
Calculator restricted	1.29	2.36	2.66	3.69	121	211.77	0.000

Table 48

Standardized Mean Difference Effect Size for Calculator Allowed and Calculator Restricted Items by Frequency of Calculator Use (Never is Comparison Group)

Item's calculator designation	Frequency of calculator use			
	Daily	Weekly	Monthly	Never
Calculator allowed	0.29	0.10	0.01	0.00
Calculator restricted	0.36	0.18	0.15	0.00

Table 49

Counts and Percentage of Correct Application of Calculator Based on Calculator Appropriateness Category

		Calculator appropriateness category		
Data for all students	Total	Active	Inactive	Neutral
Responses	72,702[a]	11,310	24,228	37,164
Omitted	2,864	571	870	1,423
Not reached	5,466	307	654	4,505
Multiple response	10	2	4	4
Applied calculator	19,717	7,566	599	11,552
Did not apply calculator	44,645	2,864	22,101	19,680
% correct application	85.80[c]	68.76	93.75	35.37[b]
Data by frequency of calculator use				
Almost every day				
Applied calculator		3,695	238	5,912
Items presented		4,943	10,534	14,868
% correct application	88.29[c]	74.75	94.64	39.76[b]
1-2 a week				
Applied calculator		1,867	167	2,713
Items presented		2,817	6,214	8,313
% correct application	85.02[c]	66.28	93.51	32.64[b]
1-2 a month				
Applied calculator		1,091	97	1,573
Items presented		1,748	3,675	5,063
% correct application	82.78[c]	62.41	92.46	31.07[b]
Never or hardly ever				
Applied calculator		913	97	1,354
Items presented		1,495	3,151	4,415
% correct application	82.57[c]	61.07	92.76	30.67[b]

[a] The discrepancy of 11,282 between the N above and the N=83,984 in Table 43 is caused by the "cluster" items and is explained in the methods section. There is only *one* calculator use question per cluster item (e.g. M0732A1N), but it applies to *all* parts of the item it refers to.
[b] As the *neutral* category cannot have a meaningful "*% that correctly applied calculator*", the number shown is the percent of respondents that applied a calculator to the problem.
[c] Neutral items are not included in the total % that correctly applied calculator.

Table 50

Results of the Appropriate Calculator Use Computations by Frequency of Calculator Use Category

Frequency of calculator use	Number of students presented calculator items	Number of students in Other group	Number of students in High group	Percent of correct responses from Other group	Percent of correct responses from High group	Probability of being in High group and answering correctly
Daily	2,173	757 (34.8%)	1416 (65.1%)	52.1	72.9	.475
Weekly	1,222	561 (45.9%)	661 (54.1%)	50.5	66.0	.357
Monthly	746	399 (53.5%)	347 (46.5%)	45.5	63.9	.297
Never	653	354 (54.2%)	299 (45.8%)	47.6	61.9	.283

CHAPTER V

DISCUSSION

As an aid to the reader, this final chapter restates the research problem and reviews the major methods used in the study. The main emphasis of this chapter is to summarize the results and discuss their implications.

Summary

With the development of the electronic calculator came the debate of whether or not calculators should be used in mathematics education. Supporters claimed that calculators were a tool to help facilitate mathematical learning; opponents considered them a crutch that would artificially support the mathematically feeble.

Calculators are now commonplace in schools, but the debate continues with regard to what effect they have on students' mathematical achievement. After 30 years of debate, has the calculator actually become the technology that allows students to "learn more mathematics more deeply" (National Council of Teachers of Mathematics, 2000), or have calculators fulfilled the ominous prediction of being a "crutch" to support those who have achieved "calculator-assisted mathematical incompetence" (Escobales & Rothenberg, 1987, p. 73)?

As explained in Chapter III, this research analyzed the data from the 1996 NAEP to assess the large scale implementation of calculator use in schools. This analysis utilized quantitative methods within the causal-comparative (aka ex post facto) design in order to assess the cause-effect relationship between calculator use and achievement in mathematics. This method was selected because the NAEP data is specifically designed for secondary analysis procedures based on the presence or absence of a condition and not the experimental manipulation of the condition.

Findings

Four research questions were developed to assess the effects of policies advocating the widespread use of calculators in school classrooms. The results and discussion are presented in order of the research questions.

Question 1: How does frequency of calculator use in the classroom relate to mathematics achievement on the NAEP Mathematics Assessment?

Results from the student reported use of calculators clearly indicates that more frequent calculator use is associated with higher achievement levels as measured by the 1996 NAEP Main Mathematics Exam. These results are significant at the $p < 0.0001$ level when comparing the *daily* users to all three other calculator use categories. There was only a 2 point difference in score (out of 500 possible points) between the *weekly* and *monthly* use groups, and this difference is statistically nonsignificant ($p = 0.4472$). Those in the *never* group fare the worst, finishing 22 points behind the *daily* group ($p < 0.0001$), 10 points behind the *weekly* group ($p = 0.0001$), and 9 points behind the *monthly* group ($p = 0.0027$). These results indicate that a little calculator use is better than none, but to get the most out of the calculator it should be used on a daily basis. It would appear that the calculator, like mathematics itself, requires time to learn and must be practiced regularly in order to maintain proficiency.

The results above are supported by the results based on the teachers' reported use of calculators in their classroom. Though the differences in scores between student-reported and teacher-reported use are not an exact match, the general trend and its statistical significance is concurrent. There is one teacher-reported result that was different from the student-reported results. This is the comparison between the *monthly* and *never* groups, which had a difference of 9 points and $p = 0.0027$ from student data versus 6 points and a $p = 0.2176$ from teacher data. This difference is not statistically significant, continues to support the trend of higher scores with more frequent calculator use, and does little to weaken the NCTM's arguments for using calculators.

Comparisons between the 1996 and 2000 data show similar results for both administrations. In 2000 the *weekly*, *monthly*, and *never* groups gained 5 points while the *daily* users only gained 2, but the *daily* users still significantly outscore those who use the calculator less frequently for both years. The consistency of these between years' results provides evidence of the reliability of results based on calculator use and is consistent with research conducted over the past 30 years.

The effect-size calculations indicate that the daily use of a calculator produces a $d = 0.61$ in score when compared to those who never use a calculator. To put this in perspective, the average student in the *daily* group would finish at the 73rd percentile in the *never* group. Using Cohen's (1969) guidelines for interpretation, this effect size is between medium and large and would be "visible to the naked eye" (p. 23). The effect sizes for the *weekly* ($d = 0.29$) and *monthly* ($d = 0.24$) groups are considered small, but they are still worth noting. These results are consistent with the findings from prior meta-analytical studies that reported mean effect sizes ranging from 0.14 (Hembree, 1984) to 0.42 (Smith, 1996).

The practical significance is not nearly as dramatic as the statistical significance, but considering the effort required to properly incorporate a calculator into mathematics instruction relative to the potential gains from its use, it is hard to argue against using the calculator. The one thing that is noticeably absent in this analysis is any evidence to support the "calculator as crutch" theory. There is not a single instance of a less frequent calculator use group outperforming a more frequent use group.

Question 2: How is this relationship with achievement affected when potentially confounding variables are controlled (e.g., teacher's education and experience, student's socioeconomic status, parent's education level, etc.)?

Considering each of the factors that were controlled on an individual basis, the results show that:

1. There is essentially no difference in scores based on gender. Within both genders the trend is for higher scores with more frequent calculator use, with nearly all comparisons being statistically significant (exceptions are males *weekly* versus *monthly*, diff = 0, $p = 0.9902$, and females *weekly* versus *monthly*, diff = 4, $p = 0.1349$, and *monthly* versus *never*, diff = 4, $p = 0.0887$).

2. Socioeconomic status has a somewhat predictive result of high SES students outscoring low SES students. Within the higher SES group the trend is for higher scores with more frequent calculator use. The differences in scores are significant in all but one case, *weekly* versus *monthly* (diff = 2, $p = 0.056$). This result is consistent whether SES is inferred from national school lunch program eligibility or by the

school's Title I status. The statistical significance of comparisons is identical to the trends found in Research Question 1.

Within the lower SES group the trends continue to favor more frequent use of the calculator except for the comparison between the *weekly* and *monthly* groups, where the *monthly* outscores the *weekly* by 4 points based on national school lunch eligibility data and by 3 points based on the schools Title I status. Both of these differences are within a standard error of measure of each other and neither difference is statistically significant ($p = 0.255$ and $p = 0.5949$, respectively). The statistical significance of paired comparisons is mixed based on national school lunch eligibility (see Table 18) and nonsignificant in all comparisons based on the school's Title I status.

Based on this information it would appear that the calculator influences achievement, but its impact is more significant at higher SES levels and in limited cases may have a slightly deleterious effect within low SES levels.

It should be noted that this analysis was based on the financial aspect of SES, but SES involves far more than financial resources. Payne (1996) contended that the student's emotional and mental resources, external support systems, knowledge of hidden rules, and relationships with role models have much to do with student achievement. Basing a conclusion solely on financial criteria oversimplifies a complicated system.

3. Parent's level of education has a predictable pattern of higher student scores with higher parental levels of education. Within each of the five levels of parental education (i.e., less than high school, graduated high school, some after high school, graduated college, unknown) the trends consistently indicate that higher student scores

with more frequent calculator use. The statistical significance of score differences ranges from nonsignificant in all cases (less than high school, see Table 24) to significant in all cases for *daily* users (graduated college, see Table 27). The only exception to this pattern was the *weekly* group finishing 5 points higher than the *monthly* group for parent's level of education unknown. This difference was nonsignificant.

4. Comparing NEAP achievement levels, by definition, requires differences in scores between achievement levels, but within achievement levels this analysis reveals some noteworthy findings.

Recalling the graph from Figure 2, the below basic level shows a steady increase in score with more frequent calculator use, but the differences are only significant for the *daily* versus *never* comparison. At the basic and proficient levels the trend line is flat; the calculator has essentially no statistical or practical significance. At the advanced level the *daily* group is the only group with enough students in it to permit reliable estimates. The fact that 68% of students at the advanced level use a calculator almost every day may imply that at some point paper-and-pencil computations either acquiesce to technology or force students to labor in computation.

5. When controlling for the type of school attended (i.e., public or nonpublic), the trend continues to show that higher scores are associated with more frequent calculator use regardless of school type. These results are most significant within public schools, where the difference between the *daily* and *never* groups is 24 points ($p = 0.0000$). The same comparison for nonpublic schools is not as dramatic (diff = 14, $p = 0.0027$) and the results are nonsignificant for less than daily use. These results show

once again that the calculator users are doing better than the nonusers regardless of whether they are in the typical public school or a nonpublic school.

6. The two teacher factors that were controlled both turned out to be significant. The first factor, which dealt with the teacher's knowledge of the NCTM's *Principals and Standards for School Mathematics,* resulted in higher scores correlating with greater knowledge of the standards (see Table 37). However, using a calculator is only one aspect of multiple issues addressed in the NCTM standards.

A more direct assessment of the calculator issue can be found by asking whether or not the teacher had studied the use of calculators. The answer to this question turned out to be nonsignificant in 1996 (diff = 2, $p = 0.4041$), but in 2000 the difference of 7 points between those who had and those who had not was significant at $p = 0.0015$. The between years' difference of 4 points was significant at $p = 0.0175$. This was the one instance where the 2000 results were significantly different from the 1996 results, and it turned out to support the use of calculators.

In summation, controlling for the identified potentially confounding variables had no effect on the results initially found in Research Question 1. In nearly every case the trends consistently indicate that higher scores are associated with more frequent use of the calculator, and in the vast majority of comparisons, the differences are statistically significant. There was only one exception to this; the case where the *monthly* users outperformed the *weekly* users when controlling for SES, and this result was nonsignificant.

What is remarkable about the analyses performed in Research Question 2 is the consistency of the results. Some factors would be expected to show a difference between groups, but the within group results showed time and again that students who use a calculator more frequently will, on average, score higher than those who use it less frequently.

There was one situation where the within level control did not show a significant difference for one level but did for the other, and that factor was SES. Based on the school's Title I participation, lower SES students only show a 6 point difference (p = 0.4247) between the *daily* and *never* groups, while the higher SES students show a 22 point difference ($p < 0.0000$). Though this finding does not encourage calculator use in Title I schools, it also does not give any reason to believe that the calculator is detrimental in these situations.

Question 3: How is this relationship affected when the data are disaggregated by question type, where the calculator is allowed on some NAEP questions but not others?

The results of this analysis indicate that more frequent calculator use is associated with a higher percentage of correct responses regardless of whether the calculator is allowed or restricted, and there is no significant interaction effect between the two item types and the students' frequency of calculator use. The *daily* users, on average, answer 8% more items correctly than the *never* users on both the calculator-allowed and the calculator-restricted items. According to one expert, the 8% difference is equivalent to one grade level (J. A. Dossey, personal communication, July 11, 2004). This result continues to hold when questions are further divided by their difficulty levels, content strands, and ability levels. In all cases the percent of correct responses gradually steps down with each decreasing level of calculator use. The one notable exception to this is in the problem-solving ability level on calculator-allowed items. In this case the gradual steps take a sudden drop with the *weekly*, *monthly*, and *never* groups being roughly equivalent.

Looking at this question using rank-finish ordinal data illustrates just how dominating the *daily* users are. Out of the 173 questions analyzed the *daily* group finished first 142 times (82.1%) and only finished last 2 times (1.2%). At the other end of the spectrum, the *never* group only finished first 5 times (2.9%) and finished last 122 times (70.5%). The Friedman's analysis of these rank finishes was significant at the $p =$ 0.000 level for both calculator-allowed and calculator-restricted item types.

The practical significance of these results compared the *daily* to the *never* users and showed an effect size of $d = 0.29$ on the calculator-allowed items and $d = 0.36$ on the calculator-restricted items. These effect sizes are on the boundary between small and medium (Cohen, 1969), but considering how much it affects rank finish and the percentage of correct responses, it is an investment worth making. The effect sizes for the *weekly* and *monthly* groups range from $d = 0.01$ to $d = 0.18$. Differences at this level are small at best, but they do indicate that no harm is caused by using a calculator.

These results of the calculator not being detrimental to estimation and paper-and-pencil skills are consistent with prior findings (Hembree, 1984; Smith, 1996; Sutherlin, 1977; Suydam, 1979).

Question 4: How does frequency of calculator use relate to whether students recognize that it is appropriate or inappropriate to use a calculator to solve specific NAEP problems?

The results show that all four calculator use groups are equally adept at properly withholding calculator use when it is inappropriate (93% properly withheld). What they are not equal at is applying the calculator when it is appropriate and being able to come up with the correct result.

Among the *daily* users 65.1% qualify in the high group of appropriate calculator use, while within the *never* group only 45.8% qualify. Assuming that the *never* group does not know much about using a calculator, 45.8% in the *high* group is a relatively impressive result. When it comes to computing the correct answer the gap between these groups narrows, with 72.9% of the *daily* group and 61.9% of the *never* group providing a correct response to the items. The difference between these two groups really stands out when combining these two criteria. The probability that a student will appropriately apply a calculator *and* get the correct answer is as follows: *daily*, p = 0.475; *weekly*, $p = 0.357$, *monthly*, $p = .297$, and *never*, $p = .283$. When it comes to a situation that requires the calculator, the *daily* group is nearly twice as likely to answer the item correctly as the *never* group.

There is only the slightest hint that the *daily* users might tend to robotically reach for the calculator, and that evidence comes from the calculator neutral items. These items, by definition, are such that it makes no difference whether a student uses a calculator or not. The results show that the *daily* group used a calculator 40% of the time while the other groups tended to use it closer to 30% of the time. This does not imply an inappropriate use of the calculator, but does indicate that in an either/or situation the *daily* users will reach for the calculator a little more often than the other groups.

From these results it could be speculated that the *never* group has had some exposure and practice using calculators. If such treatment diffusion did not exist then this group's ability to use the calculator as well as it did on the NAEP would be truly remarkable.

Interpretation of Findings

Based on the evidence from the 1996 NAEP Eighth Grade Main Mathematics Assessment it is apparent that (a) calculator use is consistently associated with higher levels of achievement in mathematics; (b) this result is consistent when controlling for potentially confounding factors; (c) those who use calculators on a regular basis are not necessarily calculator dependant and, in fact, outperform less frequent calculator users regardless of whether the calculator is available or not; and (d) calculator users do not necessarily use the calculator inappropriately, and, if anything, could appropriately apply it more often than they currently do.

These findings coincide with the majority of research conducted on this subject over the past 30 years and are in harmony with the position statement and technology principles of the NCTM. The results provide little, if any, evidence to support the claim that using calculators will result in technology dependency and mathematical incompetence.

Implications for Practice

The appropriate use of calculators in middle school grades can increase student achievement in mathematics. Using them on a daily basis is better than a weekly or monthly basis, but any regular use is better than no use at all.

There is little reason to fear that using calculators will be detrimental to students' learning in mathematics provided that students learn the fundamentals using paper-and-pencil methods (Suydam, 1979) and that teachers comply with the NCTM's (2000) recommendation that it "not be used as a replacement for basic understanding and intuitions (but rather) to foster those intuitions and understandings." Thus, the calculator "should be used widely and responsibly with the goal of enriching students' learning of mathematics." Furthermore, "(calculators) are not a panacea. As with any teaching tool, it can be used well or poorly" (p. 25).

Teachers are ultimately the ones who must decide if, when, and how the calculator should be used: It therefore behooves them to know and understand when and how to use it. Based on the findings from this research, teachers can improve their effectiveness by increasing their familiarity and use of the NCTM *Principles and Standards for School Mathematics*, participating in workshops, seminars, and training sessions on appropriate calculator use, and implement that training with prudence and wisdom.

Finally, teachers should recognize the uniqueness of SES and its relationship to calculator use. As most schools are capable of providing a calculator to students who cannot afford one, the economic aspect of SES is not nearly as important as the other aspects of SES. The area of SES where the teachers can be a positive influential is as a role model (Payne, 1996). It is therefore imperative that teachers appropriately model the appropriate use of calculators.

Recommendations for Further Research

The first recommendation for further study deals with testing the suggestion that calculators be restricted or used on a limited basis in elementary grades (Cowdery, 1997; Hembree, 1984; Loveless, 2004). The more recent NAEP assessments are sufficiently large that sample size is no longer an issue, making it possible to replicate this study using more recent data and at all grade levels. Such replications would be informative when considering when and how to use calculators.

A second recommendation is to use the new, more capable, analytical software to analyze NAEP data. The release of the *NAEP Data Explorer* tool will allow data to be analyzed using more complex statistical procedures including regression, multiway cross tabulations, and the crossing of all verifiable data. The original release date of March 2005 has been postponed, but it should be available sometime in the early fall of 2005. Use of the *Data Explorer* will allow for expanded analysis and more detailed results. More specifically, the trends in achievement level by calculator use group (see Figure 2) are worth further investigation.

The third suggestion is to research and delineate best practices for preparing teachers to effectively use technology in the classroom. As teachers are the most influential element in the classroom it is essential that they be competent and well trained in using the tools of the trade. Studies designed to evaluate workshops, in-services, and training programs in the use of calculators would be beneficial to identifying and supporting programs that work. It would also help teachers to make informed decisions when choosing a program.

Epilogue

> *"It is unworthy of excellent men to lose hours like slaves in the labor of calculation which could safely be relegated to anyone else if machines were used"*
>
> *Gottfried Wilhelm von Leibniz*

The above statement, in essence, is true; anyone may safely use a calculator; but not everyone who does so is an "excellent" man or necessarily a competent mathematician.

The analogy of the calculator as a tool to build mathematical understanding with the saw as a tool to build a house is often used in the calculator discussion. What is overlooked in this analogy is the fact that it is not the saw that builds the house, but the carpenter.

Carpentry is far more than just cutting two-by-fours; it is about reading blueprints, recognizing relationships, knowing how to cut the wood to fit the design, and how to fasten the individual pieces and parts together into a solid structure. A skilled carpenter can produce more and better work with power tools, but without the underlying skills, he'll only make more sawdust and noise. There is also the very real danger that he will not effectively learn from his mistakes because making mistakes with power tools does not involve the significant loss of time and labor associated with manual tools. On the plus side, this author doubts that there is a single case of an unskilled carpenter cutting-off a finger using a manual saw, but there numerous cases of power saws causing significant damage to unskilled users.

To make the analogy with calculators, it is not the calculator that builds a mathematically competent person, but the teacher. Learning mathematics is far more than punching buttons; it requires reading the needs of the students, designing instruction to meet those needs, providing experiences that allow the student to discover new knowledge, and connecting the individual parts into a coherent whole. A student may be able to get more answers faster using a calculator, but without the underlying skills of arithmetic, they may only produce more button punching and incorrect answers. Even worse, the student may not realize they are computing incorrect answers; and, if it is wrong, they can just punch the buttons again because mistakes on a calculator are not nearly as time consuming and laborious to fix as mistakes using pencil and paper.

When framing a house each wall is assembled on the ground. When all four walls have been formed the first wall is raised by "live men," then held in place with a "deadman." Deadman is a term used in construction for a temporary support piece. The deadman will hold up a frame-wall, but the wall will be flimsy, knocked over by small storms, incapable of standing over time, and prohibited from any further finish work such as wiring, plumbing, sheetrock, and hung fixtures.

Continuing with the analogy, when framing mathematics the four walls of arithmetic are built on the ground. When raising a wall it may be necessary to support it so it can stand without someone there to hold it. The calculator, properly used, can provide that temporary support, but with only the support of the calculator the mathematical conceptions will be flimsy, easily compromised, incapable of long-term retention, and prohibited from further development.

With the first wall being supported by a deadman, the adjoining wall is raised and secured to the first. The joining of walls at the corners strengthens the structure and adds stability. When all four walls are properly assembled the deadman may be safely removed and the finished structure will be left strong and self-supporting.

To finish off the analogy, the calculator may be used as a "deadman," but at some point various concepts of mathematics must be fastened into a coherent structure that can stand on its own and not depend on the external support of the calculator.

This author has no illusion that the calculator in and of itself can produce a mathematician: examples of overdependence on calculators and mathematical incompetence are commonplace. But this author also sees and recognizes the good that can come from the appropriate use of calculators in helping students to acquire, understand, and become competent in mathematical knowledge.

In the end the calculator really is just a tool, and like any tool it is an inanimate object, incapable of acting for itself, and only as good as the hands that control it. This research has convinced the author that on a national level the calculator is, for the most part, being properly used in our school settings, having a positive effect on achievement, and not artificially supporting mathematical incompetence.

Concerning the question of calculator use, this author summarizes his thoughts as follows: Do not rely on the calculator to perform computations you cannot do by hand given adequate time and resources. To depend on the calculator for the answer is an abuse of technology and indicates too much reliance on external support. Using a calculator when you *can* perform the operation by hand is indicative of an "excellent man" properly using a tool to save himself the "labor of calculation."

The author wonders if Leibniz, intelligent as he was, could have possibly imagined how controversial his vision of a calculation machine would become given the invention of the hand-held electronic calculator.

REFERENCES

Benjamini, Y., & Hochberg, Y. (1994). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B, 57*(1), 289-300.

Bernhardt, V. (2003). *Data analysis for continuous school improvement* (2nd ed.). Larchmont, NY: Eye On Education.

Betts, E. A. (1937). A preliminary investigation for the value of a calculating machine for arithmetic instruction. *Education, 58*, 229-235.

Bracey, G. W. (1998). Calculations about calculators. *Phi Delta Kappan, 79*(6), 473.

Brehm, S. S., Kassin, S. M., & Fein, S. (1999). *Social psychology*. Boston: Houghton Mifflin.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Conference Board on the Mathematical Sciences and National Advisory Committee on Mathematical Education. (1975). *Overview and analysis of school mathematics grades K-12*. Washington, DC: Conference Board of the Mathematical Sciences.

Cowdery, T. W. (1997). *The great calculator debate: Concerns about calculator use in elementary schools*. Retrieved June 21, 2001, from Illinois State University, Web site: http://www.coe.ilstu.edu/jabraun/students/cowdery/LiteratureRvw.htm

Deng, H., Ferris, J., & Hombo, C. (2003, April). *A vertical scheme of bundling the NAEP booklets.* Paper presented at the National Council on Measurement in Education, Chicago.

Dick, T. B. (1988). The continuing calculator controversy. *Arithmetic Teacher, 35*(8), 37-41.

Diploudis, A. (1997). *Undusting Napier's bones*. Retrieved June 21, 2001, from Heriot-Watt University, Web site: http://www.cee.hw.ac.uk/~greg/calculators/napier/about.html

Durrance, V. R. (1964). The effect of the rotary calculator in arithmetic achievement in grades 6, 7, and 8 (Doctoral dissertation, George Peabody College for Teachers, 1964). *Dissertation Abstracts International*, *25*, 6307A.

Escobales, R. H., Jr., & Rothenberg, R. (1987). The traditional three year sequence: It should be restored and revised. *New York State Mathematics Teachers' Journal, 37*(2), 72-76.

Fehr, H. F., McMeem, G., & Dobel, M. (1957). Using hand operation computing machines in learning arithmetic. *Arithmetic Teacher, 3*, 145-150.

Fernandes, L. (2001). *A brief history of the abacus*. Retrieved June 21, 2001, from Ryerson University, Department of Electrical and Computer Engineering, Web site: http://www.ee.ryerson.ca/~elf/abacus/history.html

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction*. White Plains, NY: Longman.

Gelernter, D. (1998). Kick calculators out of class. *Readers Digest, 153*(918), 136-137.

Glass, G. V. (1978). Integrating findings: The meta-analysis of research. *Review of Research in Education, 5*, 351-379.

Goldstine, H. (1972). *The computer from Pascal to Von Neuman*. Princeton, NJ: Princeton University Press.

Hembree, R. (1984). Model for meta-analysis of research in education, with a demonstration in mathematics education: Effects of hand-held calculators (Doctoral dissertation, University of Tennessee, 1984). *Dissertation Abstracts International, 45*, 3087A.

Huck, S. (2000). *Reading statistics and research*. New York: Longman.

Hunsaker, D. (1997, November 3). Ditch the calculators. *Newsweek, 130,* 20.

Klein, D. (2001). Should we curb the use of calculators? *American Educator, 85*(6), 4.

Loveless, T. (2004, April). *Computation skills, calculators, and achievement gaps: An analysis of NAEP items.* Paper presented at the American Educational Research Association, San Diego, CA.

Michaelson, G. (1997a). *Biography of Pascal*. Retrieved June 21, 2001, from Heriot-Watt University, Web site:http://www.macs.hw.ac.uk/~greg/calculators/ pascal/About_Pascal.htm

Michaelson, G. (1997b). *Pascal's adder (Pascaline)*. Retrieved June 21, 2001, from Heriot-Watt University, Edinburgh Scotland, Web site: http://www.macs.hw.ac.uk/ ~greg/calculators/pascal/Pascaline_Frames.htm

Moursand, D. G. (1981). *Calculators in the classroom.* New York: Wiley.

Mullis, I. V. S., Dossey, J. A., Owen, E. H., & Phillips, G. W. (1991). *The state of mathematics achievement: NAEP's 1990 assessment of the nation and the trial assessment of the states* (No. 21-ST-04). Washington, DC: National Center for Education Statistics.

National Assessment Governing Board. (1994). *Mathematics framework for the 1996 and 2000 national assessment of educational progress.* Retrieved November 16, 2002, from National Assessment Governing Board, Web site: http://www.nagb.org/pubs/96-2000math/ch2.html#6

National Assessment Governing Board. (2002). *Mathematics framework for the 2003 national assessment of educational progress* (No. RN91084001). Washington, DC: Author.

National Center for Education Statistics. (1999). *The NAEP 1996 technical report* (No. NCES 1999-452). Washington, DC: U.S. Government Printing Office.

National Center for Education Statistics. (2001). *The NAEP 1998 technical report* (No. NCES 2001-509). Washington, DC: U.S. Government Printing Office.

National Center for Education Statistics. (n.d.). *The nation's report card.* Retrieved March 23, 2002, from National Center for Education Statistics, Web site: http://nces.ed.gov/nationsreportcard/about/

National Council of Teachers of Mathematics. (1974). NCTM board approves policy statement on the use of minicalculators in the mathematics classroom. *NCTM Newsletter, 11*, 3.

National Council of Teachers of Mathematics. (1999). Groping and hoping for a consensus on calculator use. *Mathematics Education Dialogues, 2*(3).

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Research Council. (1996). *Evaluation of redesigning the national assessment of educational progress.* Washington, DC: National Academy Press.

Neubauer, S. G. (1982). *The use of hand-held calculators in schools: A review.* (ERIC Document Reproduction Service No. EDED 220 272)

Parkhurst, S. (1979). *Hand-held calculators in the classroom: A review of research.* (ERIC Document Reproduction Service No. ED200416)

Payne, R. K. (1996). *A framework for understanding poverty.* Highlands, TX: aha! Process Inc.

Pomerantz, H. (1997, December 4). *The role of calculators in math education.* Paper presented at the Urban Systemic Initiative/Comprehensive Partnership for Mathematics and Science Achievement Superintendents Forum, Dallas, TX.

Rabe, R. M. (1981). *Calculators in the mathematics curriculum: Effects and changes.* (ERIC Document Reproduction Service No. ED204178)

Roberts, D. M. (1980). The impact of electronic calculators on educational performance. *Review of Research in Education, 50*, 71-78.

Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). New York: The Free Press.

Saxon, J. (1986, May 16). Classroom calculators add to math illiteracy. *Wall Street Journal,* p. 1.

Science Museum. (2001). *Babbage.* Retrieved June 21, 2001, from Web site: http://www.sciencemuseum.org.uk/on-line/babbage/index.asp

Shult, D. L. (1987). Calculator use in schools. *Arithmetic Teacher, 34*(2), 181-191.

Sigg, P. O. (1982). *The hand-held calculator: Effects on mathematical abilities and implications for curriculum change.* South Bend: Indiana University School of Education. (ERIC Document Reproduction Service No. ED218147)

Smith, B. A. (1996). A meta-analysis of outcomes from the use of calculators in mathematics education (Doctoral dissertation, Texas A & M University, 1996). *Dissertation Abstracts International, 57,787A*,

Sutherlin, W. N. (1977). The pocket calculator: Its effect on the acquisition of decimal estimation skills at intermediate grade levels (Doctoral dissertation, University of Oregon, 1977). *Dissertation Abstracts International*, *37*, 5663A.

Suydam, M. N. (1976). *Electronic hand calculators: The implications for pre-college education. Final report. Abbreviated version.* Washington DC: National Science Foundation. (ERIC Document Reproduction Service No. ED127206)

Suydam, M. N. (1978). *State of the art review on calculators: Their use in education.* Columbus, OH: Calculator Information Center. (ERIC Document Reproduction Service No. ED167426)

Suydam, M. N. (1979). *The use of calculators in pre-college education: A state-of-the-art review.* Columbus, OH: Calculator Information Center. (ERIC Document Reproduction Service No. ED171573)

Suydam, M. N. (1980). *Using calculators in pre-college education: Third annual state-of-the-art review.* Columbus, OH: Calculator Information Center. (ERIC Document Reproduction Service No. ED206454)

Suydam, M. N. (1981). *The use of calculators in pre-college education: Fourth annual state-of-the-art review*. Columbus, OH: Calculator Information Center. (ERIC Document Reproduction Service No. ED206454)

Suydam, M. N. (1982). *The use of calculators in pre-college education: Fifth annual state-of-the-art review*. Columbus, OH: Calculator Information Center. (ERIC Document Reproduction Service No. ED220273)

Weik, M. H. (1961). *The ENIAC story*. Retrieved June 22, 2001, from The US Army Research Laboratory, Web site: http://ftp.arl.mil/~mike/comphist/eniac-story.html

Williams, D. E. (1978). The effect of the use of the mini-calculator and an associated curriculum supplement on computational skills and attitudes toward arithmetic of ninth grade non-college bound students (Doctoral dissertation, Temple University, 1978). *Dissertation Abstracts International*, *39*, 6610A.

VITA

Kenneth L. Wareham

Education

Utah State University 2005	Ph.D. Research & Evaluation Methodology
Utah State University 1998	MS Instructional Technology
Utah State University 1993	Teaching Credential
Utah State University 1992	BS Geology Minor in Speech & Communication

Training

2003 *Mathematical Methods and Modeling Workshop*, Mathematical Association of American.

2000 *Seminar on the National Assessment of Educational Progress Databases,* National Center for Education Statistics, U.S. Department of Education.

2000 *AERA Institute on Statistical Analysis for Education Policy*. National Center for Education Statistics, U.S. Department of Education.

2000 *Local Systemic Change Classroom Observation Training*, Horizon Research Incorporated.

Academic Experience

2001- current: Lewis-Clark State College, Professor, Secondary Education. Teach courses in science and math methods, technology applications, assessment, statistics, and supervise/mentor student-teacher interns.

1997- current: Utah State University, Instructor, CEE 6840 *Elementary Science for Educators*.

1992-97: West Side High School, Dayton, ID. Math and Science Teacher, Science Department Chair (93-97), Gifted and Talented Director (94-97).

Professional Experience

1997-current	KLW Enterprises: Program Assessment and Evaluation.
1994-97	Idaho National Engineering Lab: Instructional Design/Development 1997 (summer), Teacher Research Associate Groundwater instructor (1995 summer), Research Hydrogeologist (1994 Summer).
1990-92	BHP Corp., Centurion Mines: Exploration Geologist.

Research

Hand-Held Calculators and Mathematics Achievement: What the 1996 NAEP 8th grade mathematics exam scores tell us. Dissertation. Utah State University.

Mathematics Association of America Professional Enhancement Programs Workshop " Mathematical Methods and Modeling for Secondary Mathematics Teacher Education" Lewis and Clark College, Portland, OR 2002.

Service

State of Idaho:

- Project Based Learning advisory, Idaho High School Redesign Project (2005).
- Foundation Member, Idaho Science, Mathematics, and Technology Coalition (2003-current).
- Chair, National Youth Science Camp Selection Committee (2003-current).
- State Invited Member of the NASA/NASSMC Linking Leaders for Systemic Improvement Program (2001-2003).
- Review Committee for the Presidential Award for Excellence in Mathematics and Science Teaching (2001-current).
- Judge, Region II INEEL Scholastic Tournament (2002-current).
- Praxis II review committee: mathematics (2001).
- Praxis II review committee: science (2002).

Lewis-Clark State College:

- LCSC Faculty Representative at the Annual Legislative Luncheon (2003, 2004).
- Advisor, Kappa Delta Pi International Honor Society in Education, Chi Kappa Chapter (2003-current).
- Student Education Association; advisor (2001-current).
- LCSC Kid's College summer program instructor: Rocket Science, 2003, 2004.
- LCSC Institutional Review Board (2003-current).
- Faculty selection committee; Department of Education (2005). Two successful hires: Michelle Doty and Cliff Matousek.
- Faculty selection committee; Department of Natural Sciences (2001, search failed), 2002 successful hire of Rachel Jameton.
- Educational Testing Service test proctor (2002-present).
- Program Assessment Committee, Education Division (2001-present).
- Secondary Education Instructional Team (2001-present).
- Kappa Delta Pi Honor Society Treasurer (2002-2003).

Utah State University

- Graduate Student Senate:
  - Finance, Travel, and Budget Committee Chair (2000-01).
  - Vice-President (1999-00).
- Geology department undergraduate student representative (1991-92).
- Science Council (1990-91).

West Side School District

- District science specialist (1993-97).
- Academic team coach (1993-97).
- District technology committee (1992-97).
- Senior class advisor (1995).
- Clock/scoreboard operator, basketball & football (1994-97).

Professional Organizations

American Educational Research Association
National Science Teachers Association
Idaho Science Teachers Association
Kappa Delta Pi International Honor Society in Education
Center for the Study of the Great Ideas

Awards and Grants

2005 LCSC Technology Advisory Committee. Co-wrote a proposal with Chris Ahlman of Social Sciences to fund purchase of SPSS software and 60 end-user licenses for one 25 station SPSS computer lab, 15 professor's computers, and 20 computers at the LCSC Coeur d'Alene camps. ($10,397)

Faculty Development Grant, Lewis-Clark State College. Funded attendance at the biennial Convocation of Kappa Delta Pi International Honor Society in Education. ($1500)

2003 *Imparting Teachers with Expertise, Appreciation, And Comprehension in History (iTEACH)*. U.S. Department of Education Teaching American History Grant, written in conjunction with Spectrum Consulting, Logan UT ($890,000).

2000 Governor's Inno vative Grant, State of Idaho. Idea for enhancing mathematics understanding at the elementary level was selected as "exe mplary and innovative." Funded purchase of materials and curriculum development. ($500)

1998 National Science Foundation/American Educational Research Association Doctoral Fellowship; Math & Science Evaluation Training Program. Selected as one of 24 students from across the nation to receive a $70,000 fellowship for research and training in program evaluation.

1997 Graduate Internship, Idaho National Engineering & Environmental Laboratory.

1996 First Place, University of Idaho Science Competition (academic team coach).
First Place, Utah State University Physics Curriculum Development contest.
State Delegate, Intermountain Junior Science Symposium.
Who's Who Among America's High School Teachers.

1995 National Teacher Research Associate (TRAC), US Department of Energy.
TRAC follow-on award grant ($500).
State delegate, Intermountain Junior Science Symposium.
NASA/Idaho Space Science Consortium Grant ($250).
3rd Place, Utah State University Physics Curriculum Development contest.

1994 Research Associate, Department of Geo-Sciences, Idaho National Engineering Lab.
Teacher Incentive Grant, Idaho Humanities Council ($600).
State delegate, Intermountain Junior Science Symposium.
Who's Who Among America's High School Teachers.

1993 State delegate, Intermountain Junior Science Symposium.
Who's Who Among America's High School Teachers.

1992 Sawtooth Environmental Science Scholarship, Boise State University ($500).
Course development grant, Idaho Science Teachers Association ($350).

1991 Distinguished Service Award, Utah State University.
Utah State University Science Council.

Professional Presentations

Hand-Held Calculators and Mathematics Achievement: What the 1996 NAEP 8th grade mathematics exam scores tell us. Keynote Address, Idaho Council of Teachers of Mathematics, Moscow, ID October 2005.

Using Mathematical Modeling in the Methods Class. Association of Mathematics Teacher Educators, San Diego, CA, January 2004.

Flight of a trajectory; A sample project from the Mathematical Methods and Modeling for Secondary Mathematics Teacher Education. Mathematical Association of America, Baltimore MD, January 2003.

A simple demonstration device for AC/DC currents. Partners in Education, Lewiston, ID October 2002.

Professional Development Resources for New and Experienced Evaluators American Evaluation Association, Honolulu HI, November 2000

Evaluation Training Programs: Is the Foundation Solid? American Evaluation Association, Orlando FL, November 1999.

Evaluation Projects

Summer at WSU–Engineering Experiences for Teachers (SWEET). Washington State University College of Engineering. Independent evaluation of program goals and achievement. 2005.

Utah Elementary CORE Academy. Utah State Office of Education. Independent evaluation of in-service teacher workshops in science and mathematics instruction for grades K-6. 2005.

Imparting Teachers with Expertise, Appreciation, And Comprehension in History (iTEACH). Contract evaluator with Spectrum Consulting, Logan UT. U.S. Department of Education Teaching American History Grant, 2003-2006.

21st Century Schools. Contract evaluator with Spectrum Consulting, Logan UT.

Learning Cycle Model of Instruction, Utah State University, 2000.

Teacher Materials for Educational Training. Reviewer of professional development materials for inservice teachers. Horizon Research Inc., Chapel Hill, NC. Materials reviewed to date: Teaching High School Science, NSF; Science Standards, ASCD; Project Earth Science, NSTA; Computers, Teachers, Peers, LEA Publishing.

Local Systemic Change Evaluation, Horizon Research Inc., Chapel Hill, NC. 2000

Council of State Science Supervisors 2000 Annual Meeting. 2000

Utah Science Core Update, Utah State Office of Education 2000.

Science Teacher Need's Assessment, Utah State Office of Education 2000.

Systemic Teacher Excellence Preparation Project, Montana State University 1999-01.

Science Workshops for Teachers, Utah State Office of Education 1999.

Member Needs Assessment, Council of State Science Supervisors 1999

7th & 8th Grade Integrated Science, Utah State University 1999.

Science Workshops for Teachers, Utah State Office of Education, 1998.